Evaluating the Bank of England

Density Forecasts of Inflation

Michael P. Clements*

Department of Economics,

University of Warwick

November 5, 2002

Abstract

We consider evaluating the UK Monetary Policy Committee's inflation density forecasts using probability integral transform goodness-of-fit tests and measures related to a decision/cost based approach. In implementing the decision-based approach a number of difficulties arise in the absence of full information on payoffs, and when the actual and forecast inflation probabilities may depend on the actions taken in setting interest rates.

Journal of Economic Literature classification: C53.

Keywords: Density forecasts, probability integral transform, decision-based evaluation, inflation.

First version: August 2002.

This version: November 2002.

*Financial support from the U.K. Economic and Social Research Council under grant L138251009 is gratefully acknowledged. Helpful comments were received from David Hendry and Ken Wallis on an earlier draft.

1 Introduction

Every quarter since August 1997 the Bank of England Inflation Report has published density forecasts of the annual rate of RPIX made by the Monetary Policy Committee. This marks an important departure from the traditional concern with the central tendency or most likely outcome of the future value of the variable, and is in line with the increasing recognition that an assessment of the degree of uncertainty surrounding a point forecast is generally indispensable. The particular form of the forecast densities emphasizes possible asymmetries between upside and downside risks, presumably because these are foremost in the minds of the Monetary Policy Committee (MPC) during their deliberations over the base rate. The Bank of England's own assessment of these forecasts has tended to focus on the closeness of some measure of the central tendency and the outcomes,¹ although Wallis (2002) shows how the Christoffersen (1998) likelihood ratio goodness-of-fit statistics for evaluating interval forecasts can be interpreted as Pearson 'chi-squared goodness-of-fit' statistics, and evaluates the Bank density forecasts within this framework.

This paper pursues two approaches to the evaluation of the MPC's forecasts. Firstly, they are evaluated as forecast densities using goodness-of-fit tests based on the probability integral transform. Tests based on the probability integral transform have recently been used to evaluate probability distributions of macroeconomic variables, such as the distributions of expected inflation from the Survey of Professional Forecasters: see Diebold, Tay and Wallis (1999b). Because the MPC forecasts provide a complete description of the expected probabilities with which future inflation will fall in any given range, a comprehensive evaluation requires an assessment of all these probabilities, and not just the most likely outcome (the mode) or the mean. Secondly, we adopt a decision/cost based assessment to the evaluation of the forecasts. The MPC is especially concerned to maintain a rate of inflation in the range 1.5 to 3.5%, so that forecasting this event accurately is of primary importance, and it may matter far less whether the probabilities assigned to inflation being in, e.g., the ranges (3.5,4] and (4.5,5] are

¹For example, along with the parameter values of the RPIX two-piece density forecasts the Bank gives one and two-year ahead average and average absolute errors for the mode and mean.

well calibrated (in the sense of being an accurate assessment of the likelihoods of these outcomes). One of the advantages of having the complete densities is that it is possible to read off the probabilities of particular events of interests, and to evaluate these, rather than evaluating the whole density. We pursue this avenue within a decision-based framework. Although there are a number of problems in implementing a decision-based approach in full, this viewpoint motivates and provides a rationale for other ways of assessing quality which can be shown to be equivalent in some circumstances. Of particular interest will be the degree of conformity between the results of the methods of evaluation we stress and those of the traditional point forecast evaluation approach.

The plan of the paper is as follows. Section 2 briefly discusses goodness-of-fit tests based on the probability integral transform, and related empirical literature. Section 3 discusses the decision-based approach, problems with its implementation in the present context, related measures of forecast quality, and possible solutions. Section 4 describes the nature of the MPC forecasts and an assessment of their quality based on the tests discussed in sections 2 and 3. Section 5 concludes.

2 Probability integral transform goodness-of-fit tests

The key idea of the probability integral transform (p.i.t.) can be traced back at least to Rosenblatt (1952), and has recently been popularised in the econometrics literature by Diebold, Hahn and Tay (1999a). Assume we have a series of n forecast densities for a random variable Y_t , made at $t - h$, which we denote by $p_{Y,t-h}(y_t)$, where $t = h + 1, \dots, n + h$. The p.i.t.s are the forecast probabilities that the random variables Y_t are no greater than the realised values, y_t :

$$z_t = \int_{-\infty}^{y_t} p_{Y,t-h}(u) du \equiv P_{Y,t-h}(y_t).$$

z_t has support on the unit interval by construction. Let $q_t(z_t)$ denote the density of z_t . If $f_{Y,t-h}(y_t)$ denotes the true density, then using the change of variable formula for $z_t = P_{Y,t-h}(y_t)$, we obtain q_t as:

$$\begin{aligned} q_t(z_t) &= f_{Y,t-h}\left(P_{Y,t-h}^{-1}(z_t)\right) \left| \frac{\partial P_{Y,t-h}^{-1}(z_t)}{\partial z_t} \right| \\ &= \frac{f_{Y,t-h}\left(P_{Y,t-h}^{-1}(z_t)\right)}{p_{Y,t-h}\left(P_{Y,t-h}^{-1}(z_t)\right)}. \end{aligned} \quad (1)$$

When the forecast density equals the true density $q_t(z_t) = 1$ for $z_t \in [0, 1]$, i.e., $z_t \sim U(0, 1)$. Even though the actual conditional densities may be changing over time (as indicated by the t subscript) provided the forecast densities match the actual densities at each t , then $z_t \sim U(0, 1)$ for each t , so that the time subscript on q_t is redundant. Moreover, when $h = 1$, so that the densities are 1-step ahead, the z_t variables are independently distributed, such that the time series $\{z_t\}_{t=1}^n$ is independently identically uniform distributed, i.e., iid $U[0, 1]$.

Evaluating the forecast densities, by assessing whether $\{z_t\}_{t=1}^n$ is iid $U[0, 1]$, involves a joint hypothesis of independence and uniformity. Independence can be assessed informally by examining correlograms of $\{z_t\}_{t=1}^n$, and of powers of this series as a check for dependence in higher moments, which would be incompatible with the independence claim, and formal tests of autocorrelation can be performed. Uniformity can be assessed by testing whether the empirical cdf of the $\{z_t\}$ is significantly different from the theoretical uniform cdf (a 45° line) using the Kolmogorov Smirnov (KS) test of whether the maximum difference between the two cdfs exceeds some critical value. Unfortunately the effect of a failure of independence on the distribution of the test for uniformity is unknown, and could be exacerbated by a small sample size. Moreover tests of autocorrelation will be affected by failure of the uniformity assumption. In view of this, graphical analyses often complement more formal tests.

Recently Berkowitz (2001) has suggested taking the inverse normal CDF transformation of the $\{z_t\}_{t=1}^n$ series, to give, say, $\{z_t^*\}_{t=1}^n$, on the grounds that more powerful tools can be applied to testing the null that the $\{z_t^*\}_{t=1}^n$ are iidN(0, 1) (for $h = 1$) compared to one of iid uniformity of the orig-

inal $\{z_t\}_{t=1}^n$ series. He proposes a one-degree of freedom test of independence against a first-order autoregressive structure, as well as a three-degree of freedom test of zero-mean, unit variance and independence. In each case the maintained assumption is that of normality, so that standard likelihood ratio tests are constructed using the gaussian likelihoods. The assumption of normality of $\{\tilde{z}_t\}_{t=1}^n$ is also amenable to testing, for example, using the Shenton and Bowman (1977) two-degree of freedom asymptotic chi-squared test or the test recommended by Doornik and Hansen (1994).

3 Decision-based approach

The importance of decision or cost based assessment of the quality of forecasts has been widely accepted for a long time. In principle, one would like to evaluate forecasts in terms of the ‘economic value’ that results from decisions made on the basis of those forecasts, where the value will typically depend on the action taken and the state of nature that eventuates. In general, decision-based forecast evaluation criteria will not coincide with traditional statistical measures of forecast accuracy, such as mean squared forecast error (MSFE) – see Pesaran and Skouras (2002, p. 245-7).² The exchanges in Clements and Hendry (1993a, 1993b) touch on the relevance of context-specific cost functions in macroeconomics, and are broadly in line with the view that “Unfortunately the complexity of situations prevents a single definitive cost measure being formulated. Thus a range of simple statistical measures have been developed to measure various aspects of forecast quality” (Encyclopaedia of Statistical Sciences, entry on Forecasting).

Whilst there is little decision-based forecast evaluation in macroeconomics, Granger and Pesaran (2000a, 2000b) note that it is common in meteorology (see, e.g., Katz and Murphy (1997)), and occurs in empirical finance (see, e.g., Leitch and Tanner (1991, 1995)). Pesaran and Skouras (2002, p. 261)

²Except in the exceptional circumstances that the cost function is quadratic, and any constraints are linear. Then, the optimal decision depends only on the conditional expectation of the variable being forecast, rather than the whole forecast density. Further, economic value, the evaluation criterion in the decision-based context, is proportional to the mean squared forecast error (MSFE) criterion. In these circumstances, the minimum MSFE *point* forecast is sufficient to generate optimal decisions and maximise economic value.

suggest that its takeup in macroeconomic forecasting will be slow because of the requirement of a full specification of the decision problem, including a mapping from forecasts to decisions, and quantification of the economic costs and benefits emanating from actions taken based on those decisions in different states of nature. Moreover, when we consider the behaviour of the government, as in the case of the MPC making decisions on interest rates, actions may affect the probabilities of outcomes. In meteorological applications, the likelihoods of different outcomes will usually be independent of agents' actions.³

Section 3.1 outlines the simple two-state, two-action decision problem. Section 3.2 then considers the issues that arise in attempting to evaluate the UK Monetary Policy Committee's (MPC) probability density forecasts of inflation in this framework. Section 3.3 considers related measures.

3.1 Two-state, two-action decision problems

As described by, e.g., Granger and Pesaran (2000a, 2000b) and Pesaran and Skouras (2002), the two possible states are called “Bad” and “Good”, where in the inflation-targetting example, the former corresponds to the outturn that inflation exceeds the target rate, and the latter to inflation (in period $t + 1$) not exceeding the target. There are two possible actions open to the decision-maker. To take action, “Yes”, indicated by $y_t = 1$, or to decline to take action, “No”, $y_t = 0$. The former corresponds to raising interest rates, the latter to cutting rates or leaving them unchanged.⁴ The states are denoted by s_{t+1} , where $s_{t+1} = 1$ indicates the Bad state materializes, and $s_{t+1} = 0$ the good state. The payoff matrix associated with this decision problem is given in table 1. In the Payoff Matrix, $U_{t+1,by}$ represents the decision maker's utility if the Bad state occurs after the Yes decision is taken, and so on.

³Whether the roads are gritted overnight will not affect the probability of temperatures falling below freezing, for example. Of course, phenomena of the sort typified by Douglas Adams' Rain God, Rob McKenna (in 'So long, and thanks for all the fish', 1984, London: Pan Books Ltd.) are possible.

⁴Formally, the two-state, two-action format is too simple for the inflation targetting example. The MPC's remit is to maintain an inflation rate between 1.5 and 3.5%. This defines the Good state. The Bad state is realised when the inflation rate is outside this range, and an explanatory open letter is required to be written to the Chancellor. Hence there are three qualitative designations of the action: raise or cut, depending on which of the thresholds might be violated, and leave unchanged. We adopt the simpler formulation, and assume that only the upper threshold is binding in practice because of the underlying aim of maintaining healthy rates of output growth.

Table 1 Payoff Matrix for a Two-State, Two-Action Decision Problem.

		<i>States</i> (s_{t+1})	
		Bad ($s_{t+1} = 1$)	Good ($s_{t+1} = 0$)
<i>Decisions</i> (y_t)	Yes ($y_t = 1$)	$U_{t+1,by}$	$U_{t+1,gy}$
	No ($y_t = 0$)	$U_{t+1,bn}$	$U_{t+1,gn}$

Given this Payoff Matrix, which maps combinations of outcomes (or realized states) and (prior) actions to economic values (U), what is the value at period t to the decision maker of a particular forecast probability of the Bad event occurring in period $t + 1$?

In the general case when the probability of the states may depend on the actions, Granger and Pesaran (2000b) define the following four conditional probabilities:

$$\Pi_{t,gy} = \Pr(s_t = 0 \mid y_{t-1} = 1, \Omega_{t-1}) \quad (2)$$

$$\Pi_{t,gn} = \Pr(s_t = 0 \mid y_{t-1} = 0, \Omega_{t-1})$$

$$\Pi_{t,by} = \Pr(s_t = 1 \mid y_{t-1} = 1, \Omega_{t-1})$$

$$\Pi_{t,bn} = \Pr(s_t = 1 \mid y_{t-1} = 0, \Omega_{t-1})$$

When it is reasonable to assume that the state outcome does not depend on the action,

$$\pi_t \equiv \Pi_{t,by} = \Pi_{t,bn} = 1 - \Pi_{t,gy} = 1 - \Pi_{t,gn}. \quad (3)$$

This assumption is made in the examples considered by Granger and Pesaran (2000a, 2000b) and Pesaran and Skouras (2002), but is clearly unrealistic when interest rates are increased precisely in order to prevent the forecast (under unchanged policy) of high inflation materialising.

In general the complete forecast density of the variable s_{t+1} will be required to carry out decision-based evaluation. Here the random variable is binary, so that $\hat{\pi}_{t+1}$ and $(1 - \hat{\pi}_{t+1})$ constitute the complete forecast density (the probabilities that $s_{t+1} = 1$ and 0), but when s_t can take on a range of values a full elucidation of the probabilities associated with each value is required. Here, $\hat{\pi}_{t+1}$ is also the forecast

of the conditional expectation of s_{t+1} , so that the point forecast and forecast density coincide for the *event* of interest.

In the general case, estimates of the four ‘action-dependent’ probabilities are required, and are denoted as $\hat{\Pi}_{t+1,gy}$ etc. Then, the expected utility of taking action ($y_t = 1$) given the set of forecast probabilities is:

$$U_{t+1,by}\hat{\Pi}_{t+1,by} + U_{t+1,gy}\hat{\Pi}_{t+1,gy}$$

and of not acting:

$$U_{t+1,bn}\hat{\Pi}_{t+1,bn} + U_{t+1,gn}\hat{\Pi}_{t+1,gn}$$

so action is taken if:

$$U_{t+1,by}\hat{\Pi}_{t+1,by} + U_{t+1,gy}\hat{\Pi}_{t+1,gy} > U_{t+1,bn}\hat{\Pi}_{t+1,bn} + U_{t+1,gn}\hat{\Pi}_{t+1,gn}.$$

When (3) holds, this simplifies to the condition that action will be taken if

$$\hat{\pi}_{t+1} > \frac{c_{t+1}}{1 + c_{t+1}} = q_{t+1}.$$

where:⁵

$$c_{t+1} = \frac{U_{t+1,gn} - U_{t+1,gy}}{U_{t+1,by} - U_{t+1,bn}} > 0$$

Thus, the link between the decision (d_t) and the forecast of the bad state ($\hat{\pi}_{t+1}$) is given by:

$$y_t^* = I(\hat{\pi}_{t+1} - q_{t+1}),$$

where $I(\cdot)$ takes the value unity when the argument is positive and zero otherwise.

The economic benefit that accrues at period $t + 1$ will depend on which state materialises and the

⁵Notice that $U_{t,by} > U_{t,bn}$, and $U_{t,gn} > U_{t,gy}$.

action taken at period t :

$$\begin{aligned}
 v_{t+1}(y_t; s_{t+1}) = & U_{t+1,by}s_{t+1}y_t \\
 & + U_{t+1,gy}(1 - s_{t+1})y_t \\
 & + U_{t+1,bn}s_{t+1}(1 - y_t) \\
 & + U_{t+1,gn}(1 - s_{t+1})(1 - y_t)
 \end{aligned}$$

Using the optimal decision rule:

$$\begin{aligned}
 v_{t+1}(y_t^*; s_{t+1}) = & U_{t+1,by}s_{t+1}y_t^* \\
 & + U_{t+1,gy}(1 - s_{t+1})y_t^* \\
 & + U_{t+1,bn}s_{t+1}(1 - y_t^*) \\
 & + U_{t+1,gn}(1 - s_{t+1})(1 - y_t^*)
 \end{aligned} \tag{4}$$

where substituting $y_t^* = I(\hat{\pi}_{t+1} - q_{t+1})$ gives v_{t+1} as a function of the forecast probability,

$$v_{t+1}(\hat{\pi}_{t+1}; s_{t+1}).$$

We can write this as:

$$v_{t+1}(\hat{\pi}_{t+1}; s_{t+1}) = a_{t+1} + b_{t+1}(s_{t+1} - q_{t+1})I(\hat{\pi}_{t+1} - q_{t+1}), \tag{5}$$

where:

$$a_{t+1} = s_{t+1}U_{t+1,bn} + (1 - s_{t+1})U_{t+1,gn}, \tag{6}$$

and:

$$b_{t+1} = U_{t+1,by} - U_{t+1,bn} + U_{t+1,gn} - U_{t+1,gy} > 0. \tag{7}$$

where the term a_{t+1} does not depend on the probability forecast estimate, $\hat{\pi}_{t+1}$.

The expected economic value of using the probability forecast $\hat{\pi}_{t+1}$ is given by

$$E[v_{t+1}(\hat{\pi}_{t+1}, s_{t+1}) | \Omega_t] = E(a_{t+1} | \Omega_t) + b_{t+1}(\pi_{t+1} - q_{t+1}) I(\hat{\pi}_{t+1} - q_{t+1}). \quad (8)$$

In our setup, the optimal decision is that $y_t = 1$ whenever the true probability of the Bad state, π_{t+1} ($\pi_{t+1} \equiv E[s_{t+1} | I_{t-1}]$), exceeds q_t , and $y_t = 0$ whenever $\pi_{t+1} < q_{t+1}$. So any forecast $\hat{\pi}_{t+1}$ that satisfies $\hat{\pi}_{t+1} > q_{t+1}$ when $\pi_{t+1} > q_{t+1}$, and $\hat{\pi}_{t+1} < q_{t+1}$ when $\pi_{t+1} < q_{t+1}$, belongs to the optimal set. But only $\hat{\pi}_{t+1} = \pi_{t+1}$, i.e., when $\hat{\pi}_{t+1}$ is estimated using the correct conditional model of s_{t+1} , has the property of not depending on q_{t+1} , and will lead to the correct decisions being made by consumers who may differ with regard to q_{t+1} .

3.2 Probabilities depend on actions

In principle, the Repo rate is set by the MPC to bring about a satisfactory profile of inflation up to the medium term, two year ahead horizon, taking into account likely developments in the economic environment.⁶ Given the delays in the effects of changes in monetary policy impacting on the economy, and especially the rate of inflation *excluding* mortgage interest payments, it seems reasonable to treat the short-horizon (say, current and next quarter) actual probabilities and forecast probabilities of the Bad state – inflation exceeding the target – as being independent of changes in the Repo rate. Then the standard two-state, two-action decision-based evaluation framework may be applicable. If we assume that from a tactical perspective, the MPC's credibility may be less harmed when inflation exceeds the target in $t + 1$ if interest rates had been raised in period t , i.e., it is better to 'be seen to be doing

⁶This is borne out by a casual reading of the MPC minutes. For example, minute 28 of the November 1997 meeting refers to the central projection at the two-year horizon exceeding the target assuming unchanged interest rates, and questions whether a 25bp rise would reduce it to around $2\frac{1}{2}\%$.

something', then the standard condition that:

$$U_{t+1,by} \equiv U_{t+1}(s_{t+1} = 1, d_t = 1) > U_{t+1,bn} \equiv U_{t+1}(s_{t+1} = 1, d_t = 0)$$

holds (as discussed in the evaluation framework of section 3.1). In addition:

$$U_{t+1,gn} \equiv U_{t+1}(s_{t+1} = 0, d_t = 0) > U_{t+1,gy} \equiv U_{t+1}(s_{t+1} = 0, d_t = 1).$$

For longer horizons, such as one-year ahead, the assumption that actual and forecast probabilities are independent of the Repo rate is not tenable and the standard decision-based framework is not applicable. The Repo rate (the decision variable) is set at the time the forecasts are made, and influences the actual (and forecast) probabilities. When the actual (and therefore also the forecast) probabilities of the states depend on the actions, as in (2), action will be taken if:

$$\hat{\Pi}_{t+1,gy} > \frac{(U_{t+1,bn} - U_{t+1,by}) + \hat{\Pi}_{t+1,gn} (U_{t+1,gn} - U_{t+1,bn})}{U_{t+1,gy} - U_{t+1,by}} \equiv q_{t+1} \left(\hat{\Pi}_{t+1,gn} \right)$$

from rearranging:

$$U_{t+1,by} \hat{\Pi}_{t+1,by} + U_{t+1,gy} \hat{\Pi}_{t+1,gy} > U_{t+1,bn} \hat{\Pi}_{t+1,bn} + U_{t+1,gn} \hat{\Pi}_{t+1,gn}$$

with $\hat{\Pi}_{t+1,by} = 1 - \hat{\Pi}_{t+1,gy}$ and $\hat{\Pi}_{t+1,bn} = 1 - \hat{\Pi}_{t+1,gn}$. The form of q_{t+1} makes explicit the dependence on $\hat{\Pi}_{t+1,gn}$, while the dependence on the 'payoffs' is implicit.⁷

We can write the optimal decision rule as:

$$y_t^* \left(\hat{\Pi}_{t+1,gy}, \hat{\Pi}_{t+1,gn} \right) = I \left(\hat{\Pi}_{t+1,gy} > q_{t+1} \left(\hat{\Pi}_{t+1,gn} \right) \right).$$

⁷Of course the payoffs are not known and may be contingent upon other political and economic factors. Note that the form they enter here makes greater information demands than when probabilities are independent of actions.

Substituting for y_t^* in (4), it is apparent that the expression for economic value requires forecast probabilities of the Good (or Bad) state both when action is taken and when it is not. We observe whether rates are raised or not, that is, whether $y_t = 1$ or 0, and if we assume that actions are optimal, we can infer whether the inequality in the indicator function above holds. Thus if $y = 1$ then the reported MPC forecast probability of inflation exceeding the target (Bad state) can be interpreted as being conditional on action being taken, i.e., the reported probability is $\hat{\Pi}_{t+1,by}$, and when $y_t = 0$, the reported⁸ forecast probability is $\hat{\Pi}_{t+1,bn}$. But this is not enough as we require both $\hat{\Pi}_{t+1,by}$ and $\hat{\Pi}_{t+1,bn}$ at each t .

3.3 Statistical measures related to decision-based economic value

Ignoring the dependence of probabilities on actions, for a set of probability forecasts and states, $t = 1, \dots, T$, the part of the average realized economic value that depends on the forecasts is given by:

$$v = \frac{1}{T} \sum_{t=1}^T b_t (s_t - q_t) I(\hat{\pi}_t - q_t). \quad (9)$$

Probability forecasts are often evaluated using the quadratic probability score (QPS) of Brier (1950) and the log probability score (LPS). These are defined as:

$$\begin{aligned} QPS &= \frac{2}{T} \sum_{t=1}^T (\hat{\pi}_t - s_t)^2 \\ LPS &= -\frac{1}{T} \sum_{t=1}^T [s_t \ln \hat{\pi}_t + (1 - s_t) \ln (1 - \hat{\pi}_t)]. \end{aligned} \quad (10)$$

The QPS is bounded between 0 and 2, with lower numbers denoting more accurate. It is of course just twice the standard MSFE measure. The MSFE (or root MSFE) is a popular measure for comparing prediction errors from point-forecasting exercises that range over the real line. As a squared measure, large mistakes attract disproportionately greater penalties than small errors. The LPS is non-negative, and

⁸More accurately, the reported 2PN from which the event probability can be calculated.

penalizes large mistakes more heavily than QPS. The LPS is the negative of the average log likelihood for the logit binary choice model.

The Kuipers score (KS) is defined as:

$$Ks = H - F$$

where H is the ‘hit rate’, the proportion of the total number of Bad states that were correctly forecast, and F is the ‘false alarm’ rate, defined as the proportion of the total number of Good states that were incorrectly forecast as being Bad states. The advantage of the Ks statistic over measures such as QPS and LPS is that always forecasting the Bad state to occur (or always forecasting the Good state) will score zero, whereas such strategies may fare well on QPS and LPS. Notice that the Ks evaluates forecasts of events (whether the Bad or Good state is forecast to occur) rather than forecasts of the probabilities of events. Given the latter, we can obtain the former using $I(\hat{\pi}_{t+1} - q_{t+1})$ in the spirit of decision-based evaluation, so that the Bad state is forecast to occur when $I(\hat{\pi}_{t+1} - q_{t+1}) = 1$. Then, H and F can be expressed as:

$$H = \frac{\sum_{t=1}^T s_t I(\hat{\pi}_t - q_t)}{\sum_{t=1}^T s_t}, \quad F = \frac{\sum_{t=1}^T (1 - s_t) I(\hat{\pi}_t - q_t)}{\sum_{t=1}^T (1 - s_t)}.$$

Granger and Pesaran (2000b) show that in special circumstances the economic value criterion is proportional to Ks:

$$v = b\bar{s}(1 - \bar{s})Ks$$

where $\bar{s} = T^{-1} \sum_{t=1}^T s_t$, the estimate of the (unconditional) probability of the Bad state. To obtain this expression, the decision problem has to be simplified by assuming that $b_t = b$, all t , $q_t = q = \bar{s}$, all t .

Finally a relationship can be established between the evaluation of market timing or directional forecast accuracy using Pesaran and Timmermann (1992, PT) statistic and the Ks. The PT statistic is based on whether the proportion of events correctly forecast differs significantly from the proportion that would have been predicted correctly under the null that predictions and realizations are independently

distributed. Translating event probability forecasts into event forecasts using $I(\hat{\pi}_{t+1} - q_{t+1})$, Granger and Pesaran (2000b) show that:

$$PT = \sqrt{T} \frac{Ks}{\sqrt{\frac{\hat{P}_\pi(1-\hat{P}_\pi)}{s(1-s)}}}$$

where $\hat{P}_\pi = T^{-1} \sum_{t=1}^T I(\hat{\pi}_t - q_t)$, the estimated probability of a forecast of the Bad state.

Leitch and Tanner (1991, 1995) argue that ‘economic forecasts’, as opposed to relatively inexpensive ARIMA time-series model forecasts, are valuable when comparisons of the two are made on the basis of directional forecast accuracy.⁹

4 The Bank of England Monetary Policy Committee Inflation Forecasts

Since August 1997, in the mid month of every quarter the Bank of England Inflation Report has contained density forecasts of RPIX inflation for the current quarter and for every quarter up to two years ahead.¹⁰ The forecasts are given analytically by the two-piece normal (2PN) distribution, and graphically by the ‘rivers of blood’ fan chart.¹¹ This is a useful way of capturing perceived asymmetries between upside and downside risks, whilst allowing probability calculations to be undertaken using the standard normal distribution. For example, given the mode, mean and standard error of the distribution, we can derive the standard errors σ_1 and σ_2 of the two normal distributions on which the 2PN is based – see, e.g., Wallis (1999, p.109, eqns. A4 and A5) – from which the probability that $X < x$ is given by:

$$\frac{2\sigma_1}{\sigma_1 + \sigma_2} \Phi\left(\frac{x - \mu}{\sigma_1}\right)$$

⁹They argue that there is a positive correlation between forecasts of interest rates that fare well in terms of directional accuracy and forecasts which generate trading profits when used in simple trading strategies to choose whether to buy or sell futures contracts. They find that economic forecasts of real & nominal GNP and the GNP deflator fare better on directional accuracy comparisons.

¹⁰Available for each quarter from 1997:3 to the present from the Bank of England web site (homepage <http://www.bankofengland.co.uk>, Excel file [rpixinternet.xls](#)).

¹¹See Britton, Fisher and Whitley (1998) and Wallis (1999) for a discussion.

for $x < \mu$, where μ is the mode of the 2PN, and by:

$$\frac{2\sigma_2}{\sigma_1 + \sigma_2} \Phi\left(\frac{x - \mu}{\sigma_2}\right) + \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2}$$

for $x > \mu$, where Φ is the standard normal cdf.

The projections in each report are based on the assumption that interest rates remain constant throughout the forecast period. The RPIX inflation rate is the annual percentage growth in quarterly RPIX (RPI excluding mortgage interest payments, ONS code CHMK). The current quarter forecasts can be viewed as 1-step ahead forecasts, and the year-ahead forecasts correspond to a five-step ahead horizon. Table 2 contains the mode, means and standard deviations of the 2PN year-ahead forecast distributions made in Aug. 1997 through to May 2001 of 1998:3 through to 2002:2 (the latest available), as well as the outcomes, the probability integral transforms of the actual values using the 2PN, and the probability forecasts of the ‘Good state’, that inflation will be below $2\frac{1}{2}\%$. Finally, the table records the change in the Repo Rate (the ‘decision variable’) at the time the forecast is made relative to the rate at the time of the previous forecast. (The value recorded for the first forecast, August 1997, is relative to the rate of 6.25 operative from May 6th, 1997). Thus, this series cumulates the changes made in the months between the quarterly forecast origins. Table 3 contains the mode, means and standard deviations of the current forecasts and the event probabilities, and the event probabilities of 1-quarter ahead (i.e., 2-step ahead) forecasts.

4.1 Results

Table 4 shows the point forecast performance of the mean of the MPC 2PN forecast densities. The bias of the year ahead forecast mean is significant at the 5% level using a one-sided t -test, matching the finding of Wallis (2002) on a shorter data set. On average the mean forecasts are around 0.16% too high one-year ahead. Current and next quarter forecast biases are small and insignificant. The third row gives the mean squared forecast error (MSFE) associated with the point forecasts. These are hard to interpret

Table 2 MPC of England One-year Ahead Inflation forecasts.

Inflation Report	Mode	Mean	Std. Dev.	Outcome	z	$1 - \pi$	Change in Repo Rate
Aug 97	1.99	2.20	0.75	2.55	0.691	0.670	0.75
Nov 97	2.19	2.84	0.61	2.53	0.370	0.349	0.25
Feb 98	2.44	2.57	0.60	2.53	0.489	0.472	0.00
May 98	2.37	2.15	0.61	2.30	0.568	0.699	0.00
Aug 98	2.86	3.00	0.60	2.17	0.076	0.205	0.25
Nov 98	2.59	2.72	0.62	2.16	0.183	0.376	-0.75
Feb 99	2.52	2.58	0.62	2.09	0.218	0.457	-1.25
May 99	2.23	2.34	0.59	2.07	0.332	0.620	-0.25
Aug 99	1.88	2.03	0.56	2.13	0.588	0.801	-0.25
Nov 99	1.84	1.79	0.55	2.11	0.718	0.904	0.50
Feb 00	2.32	2.42	0.56	1.87	0.162	0.572	0.50
May 00	2.47	2.52	0.55	2.26	0.324	0.493	0.00
Aug 00	2.48	2.48	0.54	2.38	0.426	0.515	0.00
Nov 00	2.19	2.24	0.56	1.95	0.308	0.684	0.00
Feb 01	2.09	2.04	0.55	2.37	0.722	0.798	-0.25
May 01	1.89	1.94	0.55	1.86	0.473	0.868	-0.50

These are one-year ahead forecasts, so that the forecast published in the Aug. 1997 Inflation Report, for example, relates to the third quarter of 1998. The Mode, Mean and Std. Dev are of the 2PN, and z is the probability integral transform. $1 - \pi$ is the forecast probability that the annual rate of inflation one-year hence will be less than $2\frac{1}{2}\%$.

except relative to a rival set of forecasts or a benchmark. Any one of a number of time-series models, or more structural models, could be used to generate forecasts for this purpose, but would be unlikely to pose a particularly stern test unless they adequately captured the dramatic slowdown in inflation rates to the historically low rates observed over the last ten years: see figure 1. According to the arguments in Clements and Hendry (1999) it might be thought hard to beat a ‘no change’ forecast.¹² In fact we record the MSFE which uses the unconditional mean of inflation over the forecast period as a benchmark. This is a cheat in that it uses information that would not have been available at the time the forecast was made, but is handicapped by not being able to exploit any temporal dependence in the inflation series. Whilst the current quarter performance is good judged against this benchmark, the next quarter MSFE is not much smaller than the unconditional forecast’s MSFE, and at a year ahead the MPC’s MSFE is around three times that of the benchmark. (The decline in the benchmark MSFE as the horizon increases

¹²We note in passing that for the US Atkeson and Ohanian (2001) find that no-change forecasts are more accurate on MSFE than Phillips Curve forecasts. In a study of forecasting inflation in the G7 countries Canova (2002) finds that bivariate and trivariate models are little better than univariate models.

Table 3 MPC of England Current Quarter Inflation forecasts.

Inflation Report	Mode	Mean	Std. Dev.	Outcome	z	$1 - \pi$	$1 - \pi$ From previous Report	Change in Repo Rate
Aug 97	2.65	2.69	0.15	2.81	0.798	0.094	—	0.75
Nov 97	2.60	2.73	0.12	2.80	0.735	0.000	0.6465	0.25
Feb 98	2.60	2.64	0.24	2.59	0.432	0.287	0.1080	0.00
May 98	2.83	2.74	0.24	2.94	0.793	0.165	0.3263	0.00
Aug 98	2.51	2.56	0.24	2.55	0.495	0.418	0.7482	0.25
Nov 98	2.54	2.58	0.19	2.53	0.411	0.347	0.3680	-0.75
Feb 99	2.49	2.51	0.19	2.53	0.542	0.488	0.3681	-1.25
May 99	2.48	2.51	0.18	2.30	0.113	0.492	0.4341	-0.25
Aug 99	2.31	2.35	0.17	2.17	0.142	0.814	0.5697	-0.25
Nov 99	2.20	2.19	0.17	2.16	0.428	0.974	0.7136	0.50
Feb 00	1.93	1.96	0.17	2.09	0.795	0.999	0.9338	0.50
May 00	1.88	1.89	0.17	2.07	0.852	1.00	0.9491	0.00
Aug 00	2.38	2.38	0.16	2.13	0.059	0.771	0.9733	0.00
Nov 00	2.36	2.37	0.17	2.11	0.060	0.782	0.7924	0.00
Feb 01	1.94	1.92	0.17	1.87	0.372	1.00	0.6968	-0.25
May 01	1.90	1.88	0.17	2.26	0.991	1.00	0.9878	-0.50
Aug 01	2.31	2.31	0.17	2.38	0.665	0.876	0.9905	-0.50
Nov 01	2.00	2.10	0.28	1.95	0.314	0.917	0.8821	-1.0
Feb 02	2.14	2.24	0.27	2.37	0.699	0.831	0.8040	0.00
May 02	2.02	2.02	0.26	1.86	0.272	0.968	0.8947	0.00

Most of the entries in this table relate to current quarter (one-step ahead forecasts), so that the forecast published in the Aug. 1997 Inflation Report, for example, is of 1997:3. The column headed $1 - \pi$ *From previous Report* records forecast event probabilities of the annual rate of inflation in the current quarter made one quarter ago in the previous Inflation Report, and can be interpreted as two-quarter ahead forecasts given information delays.

Table 4 Point forecast evaluation summary statistics.

Test	1-step (current qtr.)	2-step (next qtr.)	5-step (year ahead)
Bias	−0.0049	0.0034	−0.1553
Standard error of bias	0.038	0.064	0.088
MSFE	0.0267	0.0729	0.1397
Variance of inflation	0.0935	0.0851	0.0461

The variance of inflation is equivalent to the MSFE of using the unconditional mean of inflation over the forecast period as the predictor of inflation for each period.

is due to relatively high inflation observations at the beginning of the period being dropped.)

Figure 1 Annual rate of quarterly price inflation – RPI excluding mortgage interest payments.

The p.i.t.'s for the year ahead and current quarter MPC forecasts are shown in tables 2 and 3, and plotted in figure 2 as time series and histograms (along with those for the next quarter forecasts). The results of applying the testing procedures outlined in section 2 are recorded in table 5. There is no evidence against the MPC current or next quarter forecasts. The year ahead forecasts are rejected by the Berkowitz test of zero mean and unit variance transformed p.i.t.'s: see the p -value of 0.017 in the last row of table 5. This test is unaffected by any dependence in the series of $\{\tilde{\pi}_t^*\}$, which might be present because of the overlapping nature of the year ahead forecasts (and which renders the rejection on the basis of the Berkowitz test of zero mean, unit variance and independence in the penultimate row

Figure 2 Probability integral transforms of the MPC two-piece normal density forecasts of the quarterly annual inflation rate. The figures in the left column are time-series plots, those in the right column are the corresponding histograms.

unreliable) and therefore indicates that the year ahead forecast densities are deficient. This is consistent with Wallis (2002) who finds that the year-ahead density forecasts place too much probability in the upper ranges: from figure 2 it is apparent that there are no z -values in excess of around 0.75, so that too much probability mass is attributed to relatively high inflation rates. Table 5 also records the results of the tests applied to p.i.t.'s calculated for a benchmark set of forecast densities. The benchmark assumes an unconditional Gaussian forecast density fitted to the forecast period inflation data characterises the variable at each point in time, and so the same issues are pertinent as in the point forecast evaluation exercise. One notable finding is that the benchmark densities are rejected for the current quarter on the basis of correlation in the $\{z_t^*\}$ series (first Berkowitz test), indicating that the test does have power to reject a poor sequence of forecasts, while the MPC forecasts are not rejected.

Table 6 records the LPS and QPS scores for the MPC inflation forecasts of the Bad event, along with unconditional probability forecasts, for the current, next quarter, and one-year ahead horizons. At the short-horizons the MPC forecasts are clearly preferred, though $\bar{s}$ has a smaller LPS and QPS scores

Table 5 Probability integral transform-based testing of inflation density forecasts.

Test	1-step (current qtr.)		2-step (next qtr.)		5-step (year ahead)	
	BoE	Uncond.	BoE	Uncond.	BoE	Uncond.
KS test of uniformity	0.100	0.100	0.105	0.105	0.250	0.125
Bowman-Shenton normality	0.862	0.609	0.822	0.641	0.778	0.711
Doornik-Hansen normality	0.448	0.567	0.839	0.586	0.888	0.872
Berkowitz $IN(0, 1)$	0.898	0.025	0.947	0.083	0.042	0.781
Berkowitz $N(0, 1)$	0.951	1	0.835	1	0.017	1

All the test outcomes are recorded as p -values, except for the KS test, which is the test statistic value. The 5% critical values for 1, 2 and 5 step forecasts (current, next quarter and year ahead) are 0.304, 0.312 and 0.340. The first Berkowitz test is a three degree of freedom test of zero-mean, unit variance and zero first-order autocorrelation, with a maintained hypothesis of normality of the inverse-normal cdf transformation of the probability integral transforms, the second is a two degree of freedom test of zero mean and unit variance, with a maintained hypothesis of normality. The columns headed ‘BoE’ are the MPC forecasts, those headed ‘Uncond’ assume a Gaussian forecast density calibrated on the actual data over the forecast period. For the latter, the p -values of the second Berkowitz test are unity by construction.

Table 6 Comparisons based on Statistical measures, LPS & QPS.

	LPS		QPS	
	$\hat{\pi}$	$\bar{s}$	$\hat{\pi}$	$\bar{s}$
Current quarter	0.207	0.647	0.110	0.455
Next quarter	0.358	0.624	0.227	0.432
One year ahead	0.585	0.483	0.404	0.305

$\hat{\pi}$ are the MPC event probability forecasts, and $\bar{q}$ are the unconditional forecasts, defined by $\bar{s} = T^{-1} \sum_{t=1}^T s_t$.

than $\hat{\pi}$ at the one-year ahead horizon.

Table 7 records the decision-based evaluation calculations at values of q between 0.1 and 1 (in steps of 0.1). A range of values of q is used because of ignorance of the payoffs. Here we also assume $q = q$ for all t , and $b_t = b$ for all t . When $q = \bar{s} = 0.35$ (for the current quarter forecasts), economic value is proportional to the Ks (see section 3.3). For both the current and one-quarter ahead horizons the MPC forecasts have higher economic value than $\bar{s}$ for $q < 0.8$. For the year ahead forecasts the interpretation of the figures in the table as representing ‘economic value’ is more strained, but the superiority of the unconditional forecasts for $q > 2$ is clear. Intuitively, the MPC event probability forecasts generally over-state the likelihood of the Bad state (inflation above two and a half percent) relative to $\bar{s}$, which is the proportion of Bad states that actually materialises. Thus $I(\hat{\pi}_t - q_t) = 1$ in (9) for a number of observations for which $s_t - q_t < 0$, for values of q in the range 0.3 to 0.6. For $q > 0.6$, economic value

Table 7 Comparisons based on economic value.

q	Current quarter		Next quarter		Year ahead	
	$\hat{\pi}$	$\bar{s}$	$\hat{\pi}$	$\bar{s}$	$\hat{\pi}$	$\bar{s}$
0.1	0.285	0.250	0.242	0.216	0.094	0.088
0.2	0.250	0.150	0.200	0.116	0.025	0.000
0.3	0.230	0.050	0.137	0.016	-0.038	0.000
0.4	0.190	0.000	0.084	0.000	-0.075	0.000
0.5	0.150	0.000	0.079	0.000	-0.063	0.000
0.6	0.100	0.000	0.084	0.000	-0.050	0.000
0.7	0.030	0.000	0.011	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000	0.000
0.9	0.000	0.000	0.000	0.000	0.000	0.000
1.0	0.000	0.000	0.000	0.000	0.000	0.000

Columns 2 – 7 evaluate equation (9) with $b_t = b = 1$, $\forall t$ at different values of q . $\hat{\pi}$ are the MPC event probability forecasts, and $\bar{s}$ are the unconditional forecasts, defined by $\bar{s} = T^{-1} \sum_{t=1}^T s_t$.

is zero because $I(\hat{\pi}_t - q_t) = 0$ for all t .

5 Conclusion

The MPC density forecasts of inflation provide a complete description of the expected probabilities with which future inflation will fall in any given range. A statistically-based evaluation of those forecasts should therefore assess the validity of the probabilities assigned to the whole range of future values. A decision-based approach evaluates forecasts in terms of their economic value, and to the extent that this is feasible, is clearly desirable. We find the two approaches in agreement when applied to the MPC density forecasts.

This paper has addressed a number of the problems that arise in attempting to evaluate the MPC forecasts. To obtain a set of comparisons based on economic value, a number of simplifying assumptions have to be made in the absence of full information on the payoffs (e.g., $b_t = b$ and $q_t = q$ for all t). We review existing literature which shows that when these (and additional assumptions, such as $q = \bar{s}$) are made, the comparisons based on decision-based measures will give the same results as those based on commonly-used measures such as Ks and the market timing statistics. Nevertheless, the findings in this paper indicate that the MPC short-horizon event forecasts have higher economic value

than unconditional forecasts for values of $q < 0.8$ (and are otherwise equivalent). The point forecast evaluation exercise, and results based on the probability integral transform, also favour the MPC current and next quarter forecasts over suitably chosen benchmarks. Specifically, the MPC mean forecasts are preferred on MSFE, while the benchmark density forecasts p.i.t.'s are not independent, whilst those of the MPC are.

The year ahead forecasts fare poorly however evaluated. The MSFE of the mean is around three times larger than that of the benchmark. The p.i.t. approach and the QPS, LPS and economic values similarly contribute to a negative assessment. The p.i.t. indicates that the density forecasts attribute too much probability mass to relatively high rates of inflation, and relatedly, the expected probabilities that inflation exceeds $2\frac{1}{2}\%$ exceed the ex post relative frequency of this event. There is no logical necessity that the different approaches should come to the same conclusions in general as to the quality of the different horizon MPC forecasts, but in this instance there is a clear consensus.

Finally, the MPC forecast probabilities that inflation will be outside the range 1.5 to 3.5% never exceed $2\frac{1}{2}\%$ or 30% for the current and year ahead forecasts, respectively, for any observation. That the range is never exceeded in the period under study makes it difficult to construct good benchmark event probability forecasts against which to compare the MPC forecasts, because the unconditional probability is exactly zero.

References

- Atkeson, A., and Ohanian, L. (2001). Are Phillips curves useful for forecasting inflation?. *Federal Reserve Bank of Minneapolis, Quarterly Review*, **25**, 2–11. (1).
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, **19**, 465–474.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **75**, 1–3.
- Britton, E., Fisher, P., and Whitley, J. (1998). February 1998 Quarterly Bulletin. *Bank of England*, 30–37.
- Canova, F. (2002). G-7 Inflation forecasts. mimeo, Universitat Pompeu Fabra.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, **39**, 841–862.
- Clements, M. P., and Hendry, D. F. (1993a). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.
- Clements, M. P., and Hendry, D. F. (1993b). On the limitations of comparing mean squared forecast errors: A reply. *Journal of Forecasting*, **12**, 669–676.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-Stationary Economic Time Series*. Cambridge, Mass.: MIT Press. The Zeuthen Lectures on Economic Forecasting.
- Diebold, F. X., Hahn, J. Y., and Tay, A. S. (1999a). Multivariate density forecast evaluation and calibration in financial risk management: High frequency returns on foreign exchange. *Review of Economics and Statistics*, **81**, 661–673.
- Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999b). Evaluating density forecasts of inflation: The Survey of Professional Forecasters. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger*, pp. 76–90. Oxford: Oxford University Press.

- Doornik, J. A., and Hansen, H. (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.
- Granger, C. W. J., and Pesaran, M. H. (2000a). A decision-based approach to forecast evaluation. In Chan, W. S., Li, W. K., and Tong, H. (eds.), *Statistics and Finance: An Interface*: London: Imperial College Press.
- Granger, C. W. J., and Pesaran, M. H. (2000b). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, **19**, 537–560.
- Katz, R. W., and Murphy, A. H. (eds.)(1997). *Economic Value of Weather and Climate Forecasts*. Cambridge: Cambridge University Press.
- Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review*, **81**, 580–590.
- Leitch, G., and Tanner, J. E. (1995). Professional economic forecasts: are they worth their costs?. *Journal of Forecasting*, **14**, 143–157.
- Pesaran, M. H., and Skouras, S. (2002). Decision-based methods for forecast evaluation. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 241–267. Oxford: Blackwells.
- Pesaran, M. H., and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Econometric Statistics*, **10**, 461–465.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, **23**, 470–472.
- Shenton, L. R., and Bowman, K. O. (1977). A bivariate model for the distribution of $\sqrt{b_1}$ and b_2 . *Journal of the American Statistical Association*, **72**, 206–211.
- Wallis, K. F. (1999). Asymmetric density forecasts of inflation and the Bank of England's fan chart. *National Institute Economic Review*, 106–112.

Wallis, K. F. (2002). Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting*. Forthcoming.