Assessing the MPC's fan charts

By Rob Elder of the Bank's Inflation Report and Bulletin Division, George Kapetanios of the Bank's Conjunctural Assessment and Projections Division and Tim Taylor and Tony Yates of the Bank's Monetary Assessment and Strategy Division.

The MPC places considerable weight on its economic forecasts when setting monetary policy. But there is inevitably uncertainty around the outlook for the economy, and to communicate this, the MPC publishes its projections as fan charts. This article discusses some of the issues that must be taken into account when assessing those fan charts, it reports a range of formal and informal tests of various aspects of the MPC's fan charts, and it discusses developments in the economy that may have pushed outturns away from the MPC's central projections. With only six years of fan chart projections that can be compared with outturns, the sample is too small to draw strong conclusions. But to date, at most forecast horizons, inflation and output growth outcomes have been dispersed broadly in line with the MPC's fan chart bands. That suggests that the fan charts gave a reasonably good guide to the probabilities and risks facing the MPC.

Introduction

The Monetary Policy Committee (MPC) was created by the Government in May 1997 to set interest rates for the United Kingdom. Between May 1997 and December 2003, the MPC's target was to stabilise RPIX inflation at 2.5%. Since then its target has been for CPI inflation to be 2%. There is typically a lag between changing interest rates and affecting the economy. The Committee therefore places considerable weight on the outlook for GDP growth and inflation,(1) and each quarter the MPC publishes its forecasts for growth and inflation in the *Inflation Report*.

Each quarter, the MPC assesses the extent to which its previous forecast appears to be on track. So an evaluation of short-term forecast performance is an integral part of the forecast process. And each year the MPC publishes a box in the August *Inflation Report* that assesses its forecasting record by reporting various measures of forecast accuracy. From time to time, external commentators have also made assessments of the MPC's forecasting performance. Recent examples include Pagan (2003) in a report commissioned by the Bank of England Court, the House of Lords Select Committee on Economic Affairs (2004), Wallis (2004) and Allen and Mills (2005).

This article expands on the analysis behind the boxes in the *Inflation Report* by considering a wider range of tests of forecast accuracy; looking at specific reasons that may have caused outturns to differ from the MPC's central projections at particular times; and reporting how the MPC has adapted its views in the light of those developments.

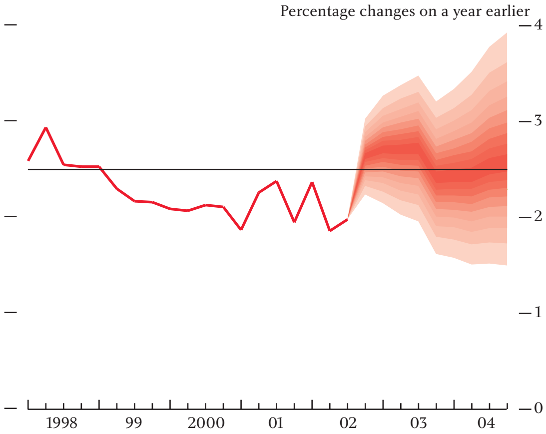
The fan charts

The MPC publishes its forecasts of inflation and output growth as probability distributions — so-called 'fan charts' — rather than as single point forecasts. The fans emphasise the inevitable uncertainty around the outlook for the economy. That could reflect uncertainty about the future economic environment: for example the outlook for world GDP, the sterling exchange rate, and other asset prices. It could also reflect uncertainty about the structure of the UK economy: for example, on the sensitivity of consumption to house price inflation or of consumer price inflation to demand growth.

Chart 1 shows an example of a fan chart for RPIX inflation. Fan charts depict the MPC's judgement of the probability of various outcomes for RPIX inflation in the future. The bands should be interpreted as follows: if economic circumstances at the start of the fan chart

(1) See Bean and Jenkinson (2001) and also the box on page 67 of the November 2000 *Inflation Report*.

Chart 1
An example of an RPIX fan chart



were to prevail on 100 occasions, inflation over the subsequent two years would be expected to lie within the darkest central band on only 10 of those occasions. Outturns of inflation are also expected to lie within each pair of the lighter red areas on 10 occasions. Consequently, inflation is expected to lie somewhere within the entire fan chart on 90 out of 100 occasions. The bands widen as the time horizon is extended, indicating the increasing uncertainty about outcomes.(1)

The shape of a fan chart reflects three judgements by the MPC about the future path of inflation (or GDP):

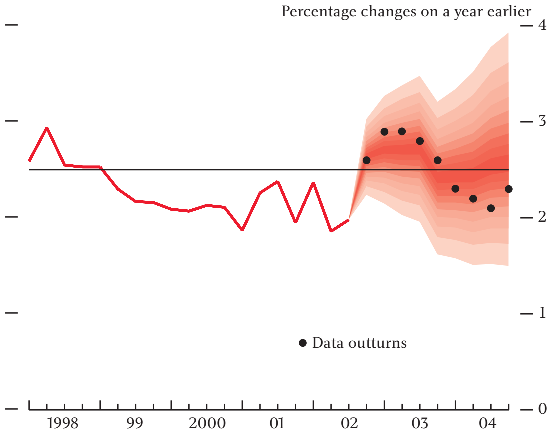
- The central projection, or the single most likely path (also called the mode(2)), which determines the profile of the central darkest band.
- The degree of uncertainty, which determines the width of the fan charts.
- And whether the fans are symmetrical or skewed, which determines the position of the mean,(2) relative to the central projection.

All of the assumptions underlying each fan chart are published on the Bank of England website.(3)

Assessing the fan charts: conceptual issues

To assess whether the MPC's fan charts have accurately described the uncertainty that it faced, we compare outturns against the probability bands (as in the example in Chart 2). By collecting this information from each fan chart, we can ask whether 10% of the outturns did actually lie in the fan chart central bands, and in each pair of outer bands.

Chart 2
Outturns relative to fan chart bands(a)



(a) The November 2002 fan chart assuming market interest rates.

The Committee publishes its fan chart projections under both the assumption of constant official interest rates and the assumption that official interest rates follow a path implied by market interest rates.(4) Sometimes, the two paths for interest rates will be similar. But when official interest rates are unusually high or low, the assumption that they will remain unchanged over the forecast period becomes less plausible. In those circumstances, assuming that interest rates follow the path implied by market expectations is likely to provide a more helpful benchmark, although there may be times when market participants hold a different view about economic prospects, and thus about the likely future course of interest rates, than the MPC. When assessing forecasting performance, it seems appropriate to focus on the fan chart based on the most plausible interest rate assumptions. So in this article, we focus on the fan charts based on the market expected path of interest rates.

There are some other important issues that must be borne in mind when considering tests of the fan charts. To date, for two year ahead projections, only 22 RPIX fan charts can be compared with data outturns (those published between February 1998 and May 2003).(5)

(1) See the box on pages 48–49 of the May 2002 *Inflation Report*, for a fuller description of the fan chart.
(2) See the box on page 332 for examples of the mode and the mean.
(3) www.bankofengland.co.uk/publications/inflationreport/irprobab.htm.
(4) See the box on pages 42–43 of the August 2004 *Inflation Report* for further discussion of this issue.
(5) This article uses RPIX data up to 2005 Q2 and GDP growth data from the 2005 *Blue Book* which goes up to 2005 Q1. So the most recent fan chart that can be compared with outturns two years ahead is from the May 2003 *Inflation Report*. The sample size is a little larger for shorter horizon forecasts. The MPC started forecasting CPI inflation in February 2004, but we do not cover CPI projections here, because to date there is not even one CPI fan chart that can be compared with outturns at the two year ahead forecast horizon.

Note that here, we do not attempt an assessment of CPI forecasts.

With such a small sample of fan charts, it is difficult to distinguish between forecasting ability and luck. Furthermore, those fan charts are not independent of one another because they overlap. Consider two projections that look two years ahead and are made one quarter apart. They will mostly cover the same time period, and so will largely be tested against the same outturns. That is likely to generate serially correlated results. If, for example, outturns were higher than the central projection from one fan chart, that would also be likely to be true of the other fan chart. As we discuss later in this article (see page 334), the combination of small samples and serial correlation can cause tests of forecasts to be misleading. So it is not wise to draw firm conclusions from informal tests (such as looking at charts), or even from more formal statistical tests, if they do not take these factors into account.

Assessing the fan charts: results

Informal analysis

Past *Inflation Report* boxes on the MPC's forecast record reported the frequency with which outturns fell in the central 30% and 50% bands of published fan charts (Table A). If the sample were large enough, and the fan charts accurately depicted the likely dispersion of outturns, then we would expect half of the outturns to lie in the central 50% bands, and 30% to lie in the central 30% bands. The actual proportions have been reasonably close to what was expected for GDP growth, with the exception that at two years ahead, a larger proportion of outturns fell in the central 50% bands (Table A). But for inflation, there were more outturns in the central bands than expected. So, this raises the question of whether the inflation fan charts might have been too wide.

Charts 3 and 4 depict similar information to Table A, on how outturns have compared with the fan chart probability bands. But the charts contain considerably more information than Table A, because they cover all forecast horizons, and all probability bands. Each dot represents a data outturn and its vertical position indicates which percentile of the fan chart it fell in at a given forecast horizon. If a dot is close to the 50th percentile line, then the outturn was close to the median projection (see the box on page 332 for a definition of the median). If a dot is above 95 or below 5, the outturn was a long way from the median and is likely to have fallen outside the visible 90% probability bands. If the MPC's fan charts have accurately depicted the true probabilities, and the samples were sufficiently large, we would expect the dots to be evenly dispersed

Table A
The dispersion of outturns relative to fan chart probability bands

	Number of outturns	Number in central 30% bands	Number in central 50% bands
RPIX inflation			
One year ahead	24	11 (46%)	15 (63%)
Two years ahead	22	8 (36%)	18 (82%)
Annual GDP growth			
One year ahead	25	8 (32%)	13 (52%)
Two years ahead	21	5 (24%)	15 (71%)

Chart 3
RPIX inflation outturns relative to fan chart probability distributions(a)

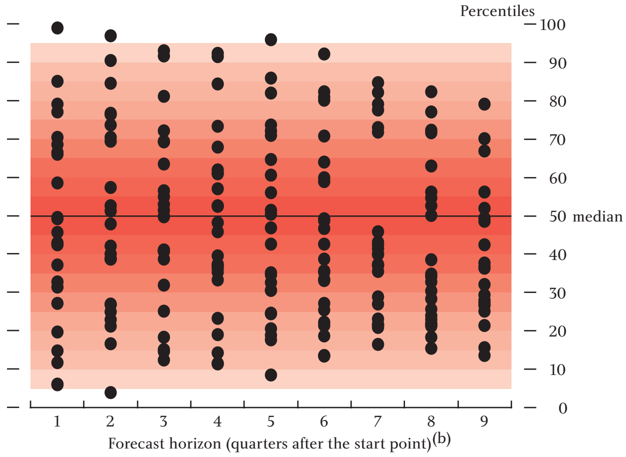
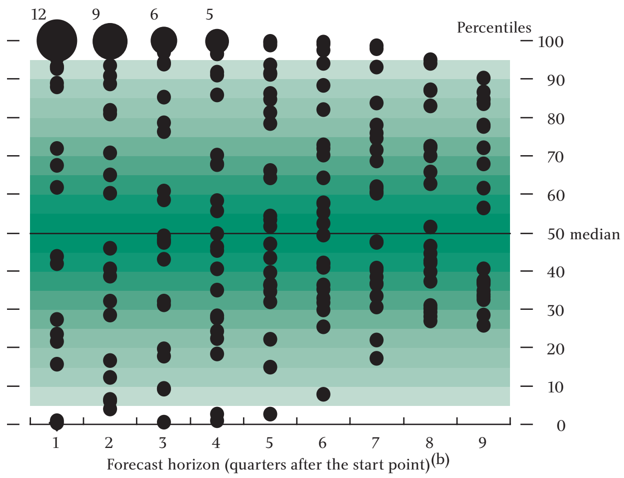


Chart 4
GDP outturns relative to fan chart probability distributions(a)(c)



(a) For forecasts made between February 1998 and November 2003 for RPIX and February 1998 to February 2005 for GDP.
(b) Data for the current quarter are not available when the forecast is made. So a 'one quarter ahead' forecast (horizon 1) is a forecast of the current quarter. For example the one quarter ahead forecast in August 2005 is for the outturn in 2005 Q3.
(c) For Chart 4, outturns at 1 to 4 quarters ahead, there are at least five outturns clustered above the 98th percentile in each case. These are shown by the larger dots in the chart; the numbers next to the dots indicate how many outturns there are.

vertically across all percentiles, for each forecast horizon.

A visual inspection suggests that in general the dots have been reasonably evenly dispersed at most forecast horizons. In broad terms, outturns have been evenly divided above and below the 50 line, with around 40% of RPIX outturns and 60% of GDP outturns above the median. This suggests that the MPC's central bands and skew have been reasonably accurate. But, for both RPIX and GDP fan charts, the dots appear clustered towards the centre for forecasts seven, eight and nine quarters ahead. That means outturns were less dispersed than the fan chart bands implied at long horizons. Furthermore, for GDP growth at short horizons, a large proportion of outturns were well above the median and clustered above the 90th percentile. That indicates both that the near-term central projections of GDP growth may have been too low, and that the fan charts may have been too narrow.

Formal statistical tests

We now turn to formal tests of the MPC's fan charts. These examine whether a set of data is likely to have been drawn from a specific distribution. The first test is a Kolmogorov-Smirnov (KS) test.[1] One problem with the KS test is that it is not very powerful in small samples. And it also takes no account of the degree of interdependence between observations. So the results must be treated with caution. The second test we use is an extension of a test first suggested by Berkowitz (2001). It is thought to be more powerful if the sample size is small, and it allows for dependence of forecast distributions over time. Nevertheless, the results must be treated with caution, as the adjustment for interdependence of fan charts is only an approximation of the expected time dependence. Appendix A describes the KS and Berkowitz tests in more detail.

Table B reports so-called 'p-values' from the two tests. They indicate how close the distribution of the outturns was to that implied by the fan charts. P-values lie between zero and one. The closer a p-value is to zero, the less likely it is that outturns were distributed in line with the fan charts. Following normal conventions, one can say with 95% confidence that there is significant evidence that the outturns were distributed differently to the fan charts if the p-values are 0.05 or below. In Table B, cells are not shaded when that is the case. But if the p-values are greater than 0.05 they are shaded orange. So shading indicates when there is not significant evidence against the fan charts. Throughout this article, when we report p-values, we have shaded them orange to indicate when the MPC's forecasts do not fail the specific test, at 95% confidence levels.

Table B
Density tests of the fan charts[(a)]

Horizon	RPIX		GDP	
	KS	Berkowitz	KS	Berkowitz
1	0.70	0.85	0.00	0.00
2	0.66	0.30	0.00	0.00
3	0.62	0.07	0.01	0.00
4	0.47	0.11	0.15	0.01
5	0.68	0.09	0.24	0.69
6	0.68	0.03	0.20	0.87
7	0.02	0.01	0.20	0.30
8	0.11	0.00	0.19	0.07
9	0.03	0.00	0.16	0.02

(a) For RPIX forecasts made between February 1998 and November 2003 and GDP forecasts made between February 1998 and February 2005. Numbers in the tables are p-values from the tests. Cells have been shaded where the p-value exceeds 5%.

The results in Table B are broadly consistent with the conclusions drawn from Charts 3 and 4. For horizons up to five quarters ahead, the distribution of RPIX outturns has been consistent with the fan charts, suggesting they have given a reasonable guide to eventual outturns.

But there is quite strong evidence that, for short horizon forecasts of GDP, the distribution of outturns was different to the fan chart distributions. In the section on GDP mean projections, we show evidence that the principal cause of the discrepancy between the short horizon GDP fan charts and the dispersion of outturns was the tendency for the published GDP data to be revised. As a result of this analysis, which was summarised in a box in the August 2005 *Inflation Report*,[2] the GDP fan charts have been widened at short horizons, which should mitigate this problem.

For long forecast horizons there is some indication that both GDP and RPIX outturns came from different distributions to the fan charts. That probably reflects outturns being more concentrated than implied by the width of the fan charts. This issue is discussed further in the next section.

Various researchers are exploring better ways to analyse forecasts of probability distributions. The tests we report here should not, therefore, be thought of as a final conclusive assessment. It is possible that different tests could generate quite different conclusions,

(1) For a textbook account of this test see Kendall and Stuart (1979).
(2) 'The MPC's forecasting record' on pages 40–41.

particularly if applied to different samples. But we take some comfort from the fact that the formal tests lead to similar conclusions as the informal analysis.

The formal tests are joint tests of the central projection, the degree of uncertainty and the skew. They do not distinguish between differences in distribution caused by the fan charts being too narrow, too wide, or centred around the wrong central profile. In the remainder of this article we focus on the individual assumptions underlying the fan charts. First, we consider the MPC's assessment of uncertainty, as the informal and formal tests of the fan charts suggest that the fan charts may have been too wide at long horizons. Then we turn to the MPC's projections of mean GDP growth and inflation.

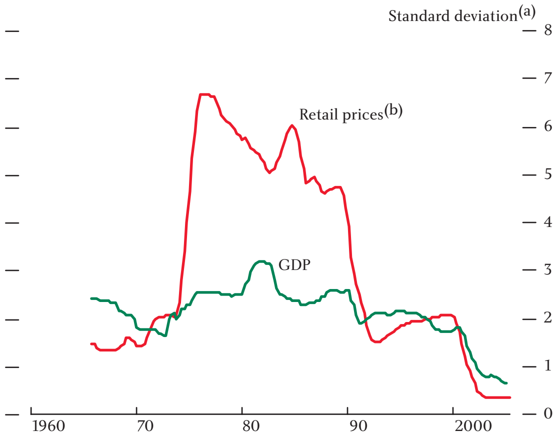
The width of the fan charts

The width of the fan charts depicts the MPC's assessment of the degree of uncertainty that it faces. Fan charts widen as the forecast period extends, reflecting the increased probability that some unforeseen event could push inflation or output growth away from the central projection. As a starting point, the width of the fan chart is based on the actual dispersion of outturns around the Bank of England/MPC forecasts over the preceding ten years. The MPC then judges whether uncertainty looking forward is likely to be greater or less than that past experience, and modifies the fan charts accordingly. As an example of such an adjustment, in February 2003 the MPC judged that the threat of a military conflict with Iraq added substantially to the risks facing the UK economy and so it temporarily widened the fan charts.

As discussed above, outturns between 2000 and 2005 tended to be closer to the MPC's two year ahead central projections than implied by the fan chart bands. Some commentators including Clements (2004), Mitchell and Hall (2005) and Wallis (2004) have highlighted this fact. Separately, Cogley *et al* (2004) generated a fan chart of UK inflation using different methods to those of the MPC. For a projection made in 2002 Q4, their fan chart was about half as wide as the equivalent chart of the MPC. So some commentators have suggested that the MPC's fan charts may have been too wide.

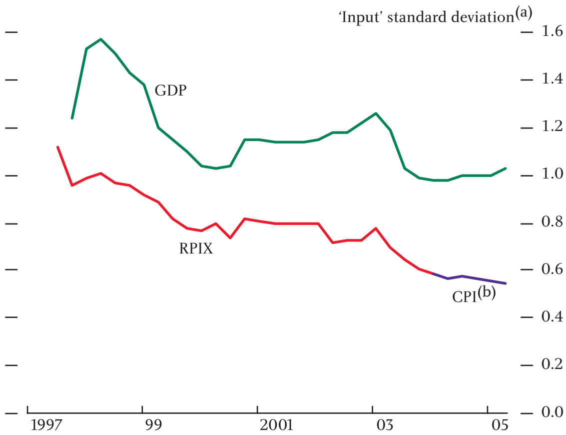
The variability of inflation and output growth has fallen substantially since the mid-1990s (Chart 5). Reflecting that, the dispersion of outturns around the Bank of England and MPC central projections has tended to fall over time. Because the MPC uses a rolling ten-year average of past forecasting experience to inform its judgement of uncertainty, the increased stability in outcomes has been translated into narrower fan charts: the width of the inflation fan chart has almost halved since 1998 (Chart 6).

Chart 5
Variability of GDP and inflation



(a) Of four-quarter rates of increase over the previous ten years.
(b) Consumption deflator between 1956–75, RPIX thereafter.

Chart 6
The uncertainty parameter in successive fan charts



(a) At the nine quarter ahead horizon. This is the parameter which the MPC chooses each quarter to set the width of the fan chart. It is σ in the appendix of Britton *et al* (1998).
(b) In the absence of information on errors forecasting CPI inflation, judgements about the width of the CPI fan charts were based on RPIX forecast errors.

The widths of the most recently published inflation fan charts are broadly in line with the variability of recent forecast errors. The standard deviation of the May 2005 fan chart was similar to the actual standard deviation of forecast errors over the previous five and ten-year windows (Table C). That suggests the MPC's judgemental adjustment to the width of the inflation fan chart in May 2005 was relatively small. And that the width of the fan chart was not overly sensitive to the size of the window of forecast errors. For GDP fan charts, the

Table C
Fan chart standard deviations compared with past outturns

	RPIX inflation		GDP growth	
	Five quarters ahead	Nine quarters ahead	Five quarters ahead	Nine quarters ahead
May 2005 fan chart	0.4	0.5	0.7	1.0
Past outturns:[a]				
since 1985 Q1	0.7	1.1	1.5	1.3
since 1995 Q1	0.4	0.4	1.0	0.8
since 2000 Q1	0.4	0.4	0.5	0.6

(a) The standard deviation of forecast errors, based on comparing the Bank of England's/MPC's mode forecast assuming constant interest rates with outturns.

assumed standard deviation is in line with the deviation since 1995 Q1, but is higher than the deviation of errors since 2000 Q1.

Should the MPC have learned about the more stable economic environment more quickly, and narrowed the fan charts accordingly? Perhaps more important than any statistical test of past fan charts is an economic assessment of the degree of uncertainty. Although we can observe the decline in variability in inflation and GDP growth in Chart 5 we cannot yet confidently explain it. We do not yet know to what extent the greater stability reflects a permanent change in the environment, such as the new monetary policy framework; and to what extent it has reflected temporary factors that may not persist.[1]

Since its inauguration, the MPC has discussed many risks to its central projections, and the fan charts have been calibrated to reflect those discussions. Some of those risks have materialised and some have not. But just because an identified risk did not crystallise, it does not mean it should not have been incorporated in the fan charts.

To take one example, the stability of inflation in recent years may have partly reflected inflation expectations remaining firmly anchored on the inflation target. But it would not be sensible for the MPC to assume that expectations will always be so anchored. There is a risk that inflation expectations could shift, say in response to a price level shock. In setting interest rates, the MPC must always remain vigilant to such a possibility.[2]

So to conclude, there does not seem to be compelling evidence that the MPC's fan charts should have been narrower. And the width of the latest inflation fan charts is in line with past experience of the volatility of outcomes.

The central tendency of the fan charts

The starting point of the MPC's forecast process is to make an assessment of the most likely path for the economy for a given profile of interest rates (see Britton *et al* (1998)). That path is called the central projection, and corresponds to the mode of the distribution of outcomes (the single most likely outcome). The Committee then considers the balance of risks around that mode projection, and may judge the distribution to be skewed. The MPC's mean projection will then reflect both its view on the most likely path for the economy, and the balance of risks. For reasons discussed in the box on page 332, in this section we will focus exclusively on the accuracy of the MPC's mean projection. But as is made clear by the fan charts, the probability that the economy will exactly follow the MPC's mean projection — or indeed any particular path — is small.

There is a large academic literature on how to assess the accuracy of such 'point forecasts'. One measure employed in the literature is the 'forecast error', which is the difference between a point forecast and the corresponding data outturn. But note that it is misleading to refer to the gap between the eventual outturn and the MPC's mean projection as an 'error'. The fan charts make clear that there is low probability of the MPC's central projections actually occurring. Nevertheless, to align with the existing literature, we will continue to use the term 'forecast error' to describe the deviation between the mean forecast and the outturn.

Assessing the mean projections: conceptual issues

One way to assess the MPC's forecasts is to compare the characteristics of their 'forecast errors' against the following, well-established criteria:[3]

(a) The mean forecast error should be zero (which implies no bias).

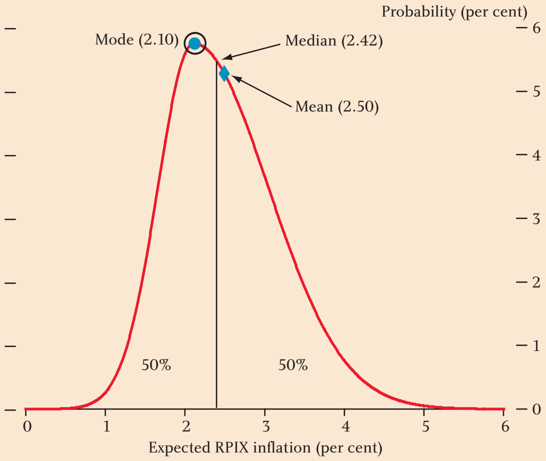
(b) It should not be possible to improve the accuracy of the forecasts by multiplying them by a constant (which we shall call 'weak efficiency').

(1) See, for example, Benati (2005), and King (2003).
(2) See, for example, King (2005).
(3) See Diebold and Lopez (1996). Note that some authors, for example Nordhaus (1987), used a different definition for weak efficiency.

Which measure of central tendency should we focus on?

For each MPC fan chart, there are potentially three measures of central tendency that could be assessed: the mean, the median or the mode.[1] The modal projection is the single most likely point, the median is the central point, with 50% probability of outturns lying on either side. And the mean is the expected outcome — the sum of all possible outcomes, weighted by their likelihood — and better reflects the entire probability distribution. These are illustrated in Chart A.

Chart A
The mean, median and mode



If the distribution in Chart A accurately described the likelihood of RPIX inflation outturns at a particular date in the future, then the expectation of the average inflation outturn would be the mean, rather than the mode or the median. Only the mean weights all possible outcomes by their probabilities.

To take an example, imagine one is offered a bet on the toss of a coin: heads wins £10, tails loses £10. Suppose the coin is weighted so that it lands on heads two thirds of the time. The most likely outcome (the mode projection) if the bet is made only once is a profit of ten pounds. But if the bet were made many times, the expected winnings would reflect the fact that a loss would occur one third of the time. So the mean forecast of profit for each play would be £6.67.

The mean projection is then the expectation of the average outturn from a large sample of observations. Indeed if the sample is large enough, and the probability distribution has been correctly specified, we would expect the average outturn to equal the mean projection. That would not necessarily be true of the mode or median projection.

Therefore, it is more appropriate to compare outturns with the MPC's projection of the mean, rather than the mode or median. Comparing the MPC's projection of the mean with outturns is akin to a joint test of the MPC's judgements on the central projection (mode), and on the balance of risks.

(1) These are made available on the Bank of England website in the week following the publication of the *Inflation Report*.

(c) Nor should it be possible systematically to use any other information, available to the forecaster at the time, to improve forecast accuracy (which we shall call 'strong efficiency').

These properties are illustrated in the box on page 333.

Regression tests

Typically, researchers test for bias and efficiency by estimating various regression equations. Let Y_t be the variable we are forecasting, and let Y_t^{t+i} represent the mean projection of Y_{t+i}, from time t the latest data point[1] (an i quarter ahead forecast). Define the i quarter ahead forecast error $e_t^{t+i} = Y_{t+i} - Y_t^{t+i}$, and let u_t be a zero-mean error term.

To test for bias, property (a), we can estimate the regression:

$$e_t^{t+i} = \alpha + u_t; \text{ unbiasedness requires } \alpha = 0 \quad \textbf{(1)}$$

For a joint test of bias and weak efficiency (b) estimate:

$$Y_{t+i} = \alpha + \beta Y_t^{t+i} + u_t; \text{ unbiasedness requires } \alpha = 0 \text{ and weak efficiency requires } \beta = 1. \quad \textbf{(2)}$$

(1) Data for the current quarter are not available when the forecast is made. So a 'one quarter ahead' forecast (Y_t^{t+1}) is a forecast of the current quarter. For example the one quarter ahead forecast in August 2005 is for the outturn in 2005 Q3.

Some illustrations of biased or inefficient forecasts

This box shows some examples of forecasts that have undesirable properties, to illustrate the regression tests.

In Chart A, both forecasts are biased. Forecast 1 is persistently too high, and forecast 2 is too low. They would fail regression test 1 as described on page 332 above.

Chart A
Examples of forecasts that are biased

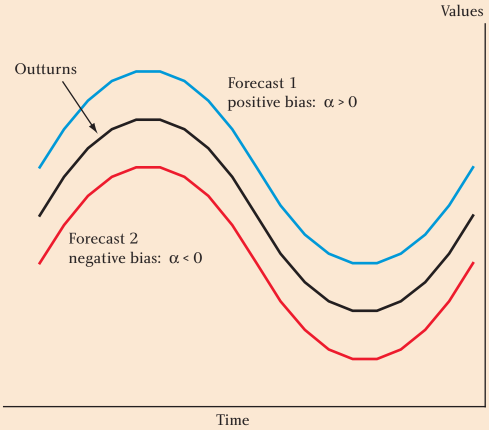
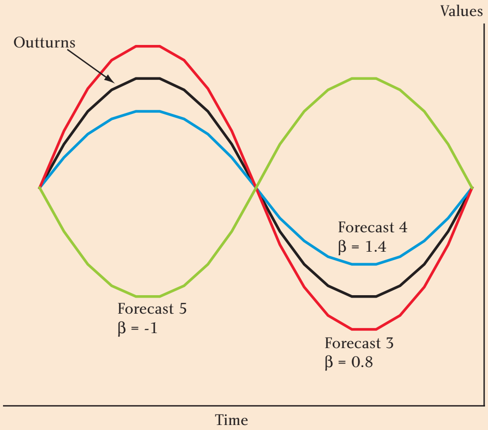


Chart B shows three forecasts, which are weakly inefficient. They have been constructed such that they are unbiased over the sample shown — they would all pass regression test 1 — but would fail regression test 2, described on page 332 above.

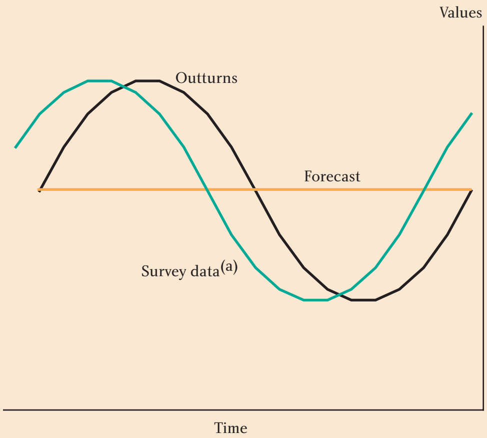
All of the forecasts in Chart B could be made more accurate if they were multiplied by a constant. Forecast 3 is too volatile, indicated by an estimated β that is less than 1. That might occur if a forecaster uses a model which puts too much emphasis on a piece of information. Forecast 4 is not volatile enough ($\beta > 1$), which might occur if the forecaster put too little weight on some information. Forecast 5 is negatively correlated with the actual outturns ($\beta = -1$). That might occur, for example, if the forecaster misinterprets some information. Chart B illustrates that even forecasts that are unbiased may not give a good guide to the future.

Chart B
Examples of forecasts that are weakly inefficient



Finally, Chart C gives an example of a forecast that is strongly inefficient. In this example, the variable is forecast to be a constant, when in fact the outturns are cyclical. By construction, the forecast is not biased, and could not be improved by being multiplied by a constant. So the forecast would pass regression tests 1 and 2. But it could clearly be improved if it took account of the survey data, which appears to lead the cyclical behaviour of the variable being forecast. So it would fail regression test 3.

Chart C
An example of a series of forecasts that are strongly inefficient



(a) Available to the forecaster before the forecast was made.

For a joint test of bias and strong efficiency (c) estimate:

$Y_{t+i} = \alpha + \beta\ Y_t^{t+i} + \lambda\, Z_t + u_t$; where Z_t is any **(3)**
information available to the forecaster at time t. Strong efficiency implies that $\alpha = 0$, $\beta = 1$, and $\lambda = 0$. In Table D, we test $\lambda = 0$ to assess the information in Z_t.

When regression tests can be misleading

Before reporting the results of regression tests of the MPC mean projections, we should note two ways that this type of regression analysis can be misleading.

Persistent errors

A run of errors in the same direction is sometimes taken as evidence of bias or inefficiency. The forecaster does not appear to have learnt from past errors. But when considering persistence in errors it is important to distinguish between forecasts of the next period (one step ahead projections), and forecasts that look many periods into the future (multi step ahead projections). It is only for one step ahead forecasts that the previous error can be observed before the next forecast is made.

Over long forecast horizons — when looking two years ahead — MPC forecast errors have tended to be persistent: there have been periods of up to two years when outturns have been consistently higher or lower than expected. But that is not necessarily evidence of poor forecasting. Indeed, we might expect to see this pattern, even if the forecasts were the best possible given the available information (see the box on page 335). Analysis of forecast errors, based on informal observation or regressions, will be misleading if it does not take sufficient account of this unavoidable persistence in errors.

Small sample size

Generally speaking, the smaller the sample size, the less powerful regression tests are. Small samples imply a greater degree of uncertainty around estimated parameters, so the results are less likely to be statistically significant. This would seem to imply that the smaller the sample, the smaller the probability of finding significant evidence that a set of forecasts is biased. But in fact the opposite can be true. For multi step ahead forecasts, with a high probability of persistence in forecast errors, the smaller the sample, the harder it can be to judge whether results are significant or not (see Appendix B).

Typically, for evaluations of two year ahead economic forecasts, researchers have used samples of at least 20 years, for example Artis (1996), Clements and Hendry (2001) and Melliss and Whittaker (1998). There have been some exceptions: an examination of the National Institute's forecast errors (Poulizac *et al* (1996)) used 13 years of data. Nevertheless, with only six years of MPC forecasts that can be compared with outturns, the sample is probably too small to draw strong conclusions.

Assessing the MPC's mean projections: results

In this section we present tests of the MPC's one step ahead and multi step ahead mean projections. It is worth noting that the GDP growth forecasts are assessed against the most recently published estimates, which for older observations may have gone through several rounds of revisions. This will prove to be important later in explaining the pattern of GDP growth forecast errors. RPIX data are not revised.

One quarter ahead mean projections

First, we report tests of the MPC's one quarter ahead projections. In forming its policy decision, the MPC places greater weight on forecasts over longer time horizons, because policy takes substantially longer than a quarter to have an effect on GDP and inflation. But one quarter ahead forecasts are still important. Each quarter the MPC monitors one quarter ahead forecast errors to assess the shocks affecting the economy. A large one quarter ahead forecast error could cause the Committee to revise its view of prospects further ahead.

The MPC has published forecasts for RPIX inflation and GDP growth since August 1997.[1] At the time of writing, there were 26 one quarter ahead projections of RPIX and 31 of GDP that could be compared with outturns. For one quarter ahead projections of RPIX inflation, as discussed above, forecast errors should not be persistent, because the forecaster observes the previous forecast error before making the next projection. So regression results should not be affected by serial correlation. And as each observation is then effectively independent, the sample may be large enough for regression results to be informative.

Regression results are reported in Table D. We report the estimated parameters behind each hypothesis test, and the p-value associated with each test in brackets. Again, p-values lie between 0 and 1, and the larger the p-value, the less evidence there is that the projections failed the test. Orange shading indicates a p-value above 5%, implying acceptable forecast performance.

The RPIX projections perform reasonably well in the regression tests. The average error is zero and is not statistically significant (test 1 in Table D). Tests for weak efficiency (test 2 in Table D) suggest that the forecasts have given a reasonably good guide to outturns — α is

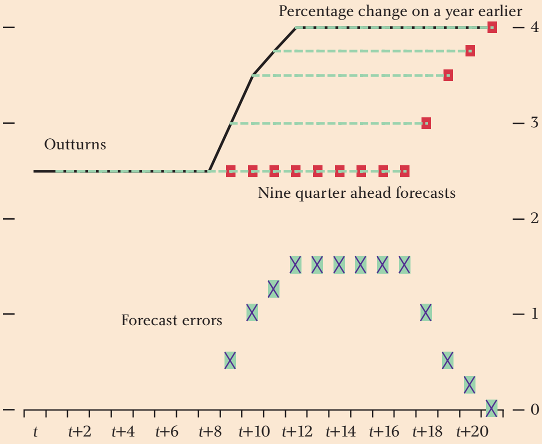
(1) The mean projection based on market interest rates was first published in February 1998. All of the MPC's projections can be downloaded from the Bank of England website.

Why multi step ahead forecasts often generate persistent forecast errors

The MPC fan charts under consideration in this article looked two years into the future, and were published each quarter. Adjacent projections overlap to a large degree. If events push outturns above the mean projection from one fan chart, it is likely that there will be similar forecast errors for other fan charts that cover that period.(1) This is illustrated in Chart A.

The red squares represent adjacent nine quarter ahead forecasts of the annual percentage change of, say, GDP. The first red square represents a forecast made at time t of growth in period $t + 9$. We assume that the best forecast is the most recent outturn. For the first nine periods, GDP growth outturns are 2.5%, so the best projection is 2.5%. Now suppose that in period $t + 9$, growth increases to 4% and stays there for several years. All the forecasts made before the shock occurs (between t and $t + 8$) turn out to be too low. So there is a long run of forecast errors, all of the same sign (the blue crosses). It is only by period $t + 9$ that the forecaster observes the shock, and can amend his forecast for growth in period $t + 18$ accordingly. So in the example, the length of the forecast period dictates the likely persistence in forecast errors, even for perfectly rational forecasts.

Chart A
Example of nine quarter ahead projections of a persistent series



(1) See Pagan, A (2003) for an algebraic example.

Table D
Regression results on one quarter ahead projections(a)

Tests:	Hypothesis	RPIX inflation	GDP growth
1) Bias	$\alpha = 0$	0.0 (0.73)	0.6 (0.01)
2) Bias and weak efficiency	$\alpha = 0$ $\beta = 1$	0.5 0.8 (0.08)	1.8 0.5 (0.05)
3) Strong efficiency(b)	$\lambda = 0$		
Z = previous forecast error		0.0 (0.88)	0.6 (0.03)
Z = previous outturn		0.0 (0.82)	0.3 (0.78)
Z = change in exchange rate		0.0 (0.90)	0.1 (0.25)
Z = import price inflation		0.0 (0.44)	–
Z = CIPS business activity		–	0.4 (0.54)

(a) For mean projections based on market expectations for interest rates. RPIX forecasts made between February 1998 and November 2003, GDP forecasts made between February 1998 and February 2005. The table cells are shaded if the p-value associated with each test (in brackets) is greater than 0.05, or in other words if at 95% confidence limits, there is no significant evidence that projections are biased or inefficient. RPIX residuals displayed no evidence of serial correlation. But it was present for GDP equations, so p-values are based on Newey West standard errors.
(b) Uses real-time GDP data, rather than latest estimates, as that is all that was available at the time the forecast was made.

not significantly different from zero and β is close to one, although the p-value is only marginally above 0.05. Four variables were included in the tests of strong efficiency (test 3 in Table D), but none of them significantly improved forecast accuracy. Tests of the GDP projections should be treated with caution as the equation residuals are serially correlated, and the Newey West correction may not be reliable given the sample size. There is significant evidence of bias, and the forecasts fail the weak efficiency test. But as we discuss below, those results can be explained by the tendency for output to be revised over time. The only evidence of strong inefficiency is that the previous forecast error (using real-time GDP) is significant. But it is not clear how much weight to put on this result given the importance of data revisions in explaining forecast errors.

Multi quarter ahead mean projections

Since 1999, each *Inflation Report* published in August has contained a box on the MPC's forecasting record. One focus of recent boxes has been to report average errors for one and two year ahead projections. These statistics are also reported on the Bank of England website.(1) As discussed above, average errors can give an indication of bias, but only if the sample size is large enough after taking account of the degree to which errors are serially

(1) www.bankofengland.co.uk/publications/inflationreport/irprobab.htm.

correlated. For one and two year ahead projections, the present sample size is probably too small to draw strong conclusions.

For the RPIX inflation projections, average errors have been close to zero, suggesting little evidence of bias (Table E). When commenting on the MPC's errors, Pagan (2003) said 'the bias is probably as small as one could reasonably expect'. Average errors forecasting GDP have been positive and further away from zero, so at least over this sample, GDP growth was on average stronger than the MPC expected.

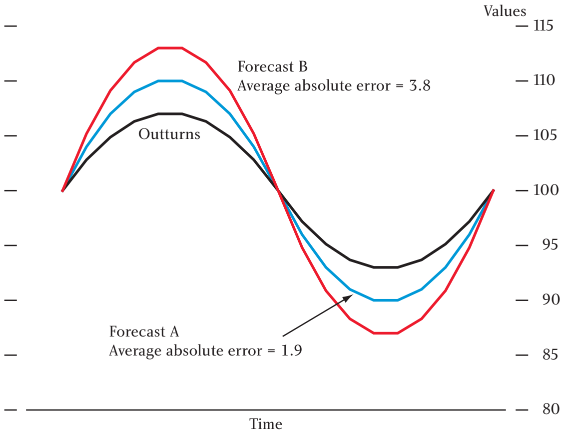
Table E
Average errors[(a)]

	RPIX inflation	GDP growth
Average errors		
One year ahead	0.0	0.5
Two years ahead	-0.3	0.3
Average absolute errors		
One year ahead	0.3	0.8
Two years ahead	0.4	0.7

(a) Calculated from mean projections based on market rates published since February 1998. For GDP projections, there are 25 observations for one year ahead projections, and 21 for two years ahead projections. For RPIX there are 24 observations for one year ahead and 22 for two years ahead.

Average errors give an indication of bias, but they cannot distinguish between an unbiased forecaster that makes large mistakes, and one that makes small mistakes. For example, in Chart 7, over the full sample, both the series of forecasts A and B have zero average forecast errors (the positive errors in the early phase are cancelled out by the negative errors in the later phase). Nevertheless, forecast A has been more accurate than forecast B. Calculating instead the average absolute errors would reveal that. The MPC average absolute errors are reported in Table E. They have been smaller for RPIX inflation than GDP growth.

Chart 7
Examples of unbiased forecasts with different absolute average errors



Forecast revisions

For multi step ahead projections, given the small sample of forecasts, and the associated serial correlation problems, the standard tests of bias and efficiency reported in Table D are not appropriate. But another test of efficiency is to examine how MPC projections of the same event (for example RPIX inflation in 2002 Q4) changed over time. Or in other words to examine the pattern of forecast revisions. Note that a forecast revision is a change in a forecast, which is quite distinct from a forecast error, which is the difference between a forecast and the eventual outturn. Unlike forecast errors revisions should not be systematically correlated over time even if forecasts overlap, see the box on page 337. Revisions can also be pooled to generate a larger sample. That allows a test of forecast efficiency for long-time horizon forecasts, even over a relatively small number of years.[1]

Table F below reports p-values from statistical tests of the degree of serial correlation, or predictability, of MPC forecast revisions. The higher the p-value, the less evidence there is of predictability, and so the less evidence of inefficiency. Following the format in Table B, cells have been shaded orange when there is no significant evidence that forecasts were inefficient at 95% confidence levels.

Table F
Tests for serial correlation in forecast revisions[(a)]

Forecast horizon[(b)]	RPIX inflation	GDP growth
6 quarters ahead	0.00	0.46
5 quarters ahead	0.01	0.10
4 quarters ahead	0.16	0.01
3 quarters ahead	0.49	0.08
2 quarters ahead	0.09	0.77
1 quarter ahead	0.45	0.88
Pooled over horizons[(c)]	0.34	0.30

(a) We test for serial correlation between revisions to a forecast of a single event by estimating the following equation:
$Y_t^{t+i} - Y_{t-1}^{t+i} = \alpha + \beta_1 (Y_{t-1}^{t+i} - Y_{t-2}^{t+i}) + \beta_2 (Y_{t-2}^{t+i} - Y_{t-3}^{t+i})$
Forecast revisions have been calculated from mean MPC projections based on market expectations of interest rates published between August 1997 and November 2004. We perform *F*-tests for one and two-period serial correlation, and report the p-values.
(b) This refers to the forecast horizon of the revision in the sequence being tested. So when $i = 1$ in the equation, it is labelled 1 quarter ahead in Table C, and so on up to 6. We stop at 6 because we require three earlier forecasts of the same event.
(c) The regression is estimated over all forecast horizons, but the constant term α was allowed to vary with the forecast horizon.

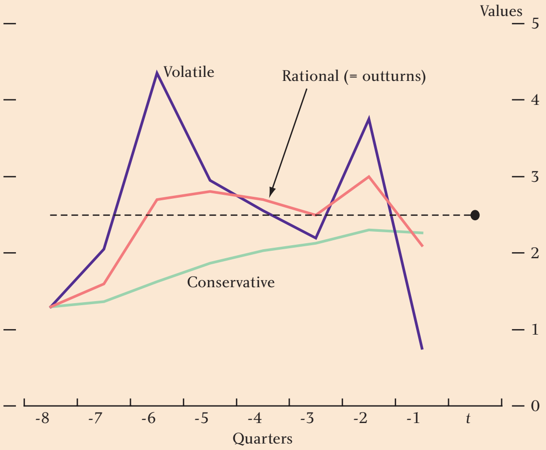
The tests indicate some evidence of correlation in revisions for long horizon forecasts of inflation, but in the majority of tests at specific horizons there is no evidence of serial correlation of forecast revisions. In the more powerful tests, when forecast revisions from all time horizons are pooled together, we do not find evidence of serial correlation. So in general, there is no

(1) See Bakhshi *et al* (2003).

Forecast revisions

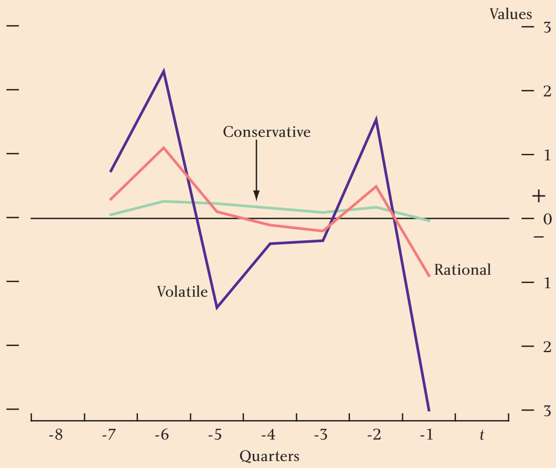
Forecast efficiency requires that revisions to forecasts of the same event should be independent over time. Otherwise, the forecaster could improve his forecast accuracy, by exploiting the predictable pattern in any forecast revisions. To illustrate this, Chart A shows examples of three types of forecaster, making successive forecasts of a single event. The rational forecaster makes optimal use of the news received each period. The conservative forecaster does not take enough account of this news, while the volatile forecaster overreacts to the news.(1)

Chart A
Successive forecasts of the outturn at period *t*



The revisions to the rational forecaster's projections are not predictable, by construction. In Chart B they appear random. This indicates that forecast accuracy could not have been improved by paying greater attention to revisions. By contrast, the conservative forecaster's revisions are predictable, because they are all in the same direction — they are positively serially correlated.

Chart B
Revisions to those forecast examples



Note: See footnote 1 below.

Similarly, the volatile forecaster's revisions are also predictable, because large revisions in one direction tend to be followed by large revisions in the opposite direction in the following period — the forecast revisions are negatively serially correlated.

Both the conservative and volatile forecaster could improve the accuracy of their forecasts by taking account of the predictability of their forecast revisions. The volatile forecaster should react less to the news, and the conservative forecaster should react more.

(1) The series being forecast is assumed to follow a random walk: $Y_t = Y_{t-1} + \varepsilon_t$. The forecasting rules in the examples are as follows:

Rational: $Y_t^{t+i} = Y_t$
Conservative: $Y_t^{t+i} = 0.2\, Y_t + 0.8\, Y_{t-1}^{t+i}$
Volatile: $Y_t^{t+i} = Y_t + 1.5\,(Y_t - Y_{t-1})$

strong evidence that MPC forecasts at most horizons could have been improved by taking into account any predictable, systematic pattern in forecast revisions. In that sense, the forecasts have been efficient.

How do MPC mean projections compare with external projections?

Another test of the MPC projections is to compare performance with equivalent projections made by other forecasters. If, over a large sample, the errors of external forecasters were clearly smaller than the MPC, that would indicate a form of strong inefficiency. Here we compare the MPC projections with forecasters surveyed by the Bank of England. Each quarter, a survey of forecasts of GDP growth and inflation is published in the *Inflation Report*.(1) The survey covers around 30 different forecasting bodies, including commercial banks, economic consultancies and academic institutions and each quarter around 20

(1) See for example the box on pages 45–46 in the February 2005 *Report*.

send in new forecasts. The forecasts are all made at roughly the same time as the MPC projection is finalised.

In Table G below, we compare the MPC average forecast error and average absolute forecast error with the equivalent measures of performance of the external forecasters. The 'pooled externals' row shows the indicators of performance of a forecast constructed by taking the average of all external forecasts. Compared with the MPC, measured bias was a little lower for this pooled forecast (indicated by a lower average error), and the size of the errors was a little smaller (as indicated by the average absolute error). That reflects the well-established result that pooling forecasts often generates a more accurate forecast than putting 100% weight on any one.

Table G
The forecast record of MPC mean projections compared with other forecasters[(a)]

	GDP		RPIX	
	Nine quarters ahead	All horizons[(b)]	Nine quarters ahead	All horizons[(b)]
Average errors				
MPC	0.3	0.5	-0.3	-0.1
External forecasters	0.3	0.4	-0.1	-0.1
Average absolute errors[(c)]				
MPC	0.7	0.7	0.4	0.3
Pooled externals	0.7	0.7	0.3	0.2
Individual externals	0.8	0.8	0.4	0.3

(a) Uses the MPC's mean projections under market expectations of interest rates published between February 1998 and November 2003 for RPIX and November 2004 for GDP. External forecasts are as surveyed by the Bank of England and published in the *Inflation Report*. Average errors are calculated on exactly the same basis for MPC projections and external forecasters.

(b) Forecasts from different horizons have been pooled, because the survey asks for a forecast at a specific point, rather than at a fixed horizon (ie RPIX inflation in 2004 Q4, rather than in two years' time). Therefore, with the exception of nine quarter ahead projections, the samples of forecasts at a given horizon are small. For example for GDP projections they are as follows: one quarter ahead to four quarter ahead, seven observations; five quarter ahead to eight quarter ahead, six observations; nine quarter ahead, 21 observations.

(c) Let $e_{i,j}$ be the forecast error of external forecaster i in relation to an event j, where each combination of forecast date and forecast horizon is considered a separate event. So for example horizon 3 and 7 forecasts at time t are separate events; and horizon 2 forecasts at time t and $t + 1$ are also separate events. Then the 'pooled external' row is calculated as $\frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{N}abs\left(\sum_{i=1}^{N}e_{i,j}\right)\right]$ and the 'individual external' row as $\frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{N}\sum_{i=1}^{N}(abs\ e_{i,j})\right]$. As described in footnote (b), the full sample for GDP comprises 73 events, while the sub-sample of nine quarters ahead forecasts has 21 observations. The samples for RPIX are 70 and 22 respectively.

The 'individual externals' row of Table G shows the average performance of individual forecasters. That removes the effect on forecast performance of pooling many different forecasts, and so in some sense is a fairer comparator for the MPC forecasts. A comparison of average absolute errors suggests that MPC forecast errors tended to be a little smaller for GDP growth, but were the same for inflation. But, just as with the other forecast tests, the sample size is too small to be conclusive.

What might have caused outturns to differ from the MPC's mean projections?

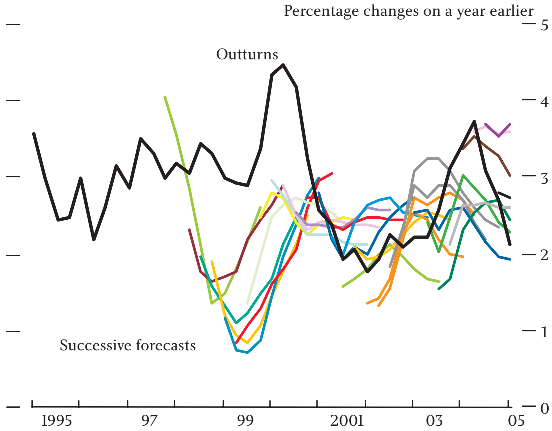
GDP mean projections

Chart 8 shows each successive mean projection made by the MPC for GDP growth. The chart highlights two periods that may be worth focusing on:

- Unexpectedly strong growth in 1999 and 2000.
- Unexpectedly weak growth in 2002.

One feature of Chart 8 is that the start point of projections made between 1998 and 2000 now appears to be too low, suggesting that growth has been revised up quite markedly since the projections were made. That is indeed the case (Chart 9). On average, the initial estimates of four-quarter GDP growth between 1998 Q3 and 2000 Q4 were 1.1 percentage points weaker than the current estimates. Furthermore, business surveys published at the time also suggested that growth was weaker than current ONS data suggest (Chart 10). The

Chart 8
GDP outturns and central projections[(a)]



(a) Mean projection based on market interest rates.

Chart 9
GDP outturns: initial and latest estimates

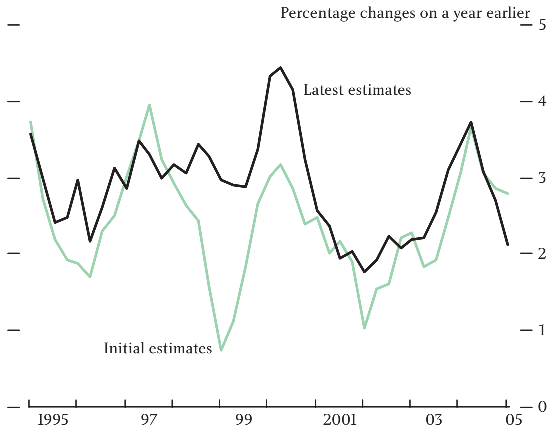
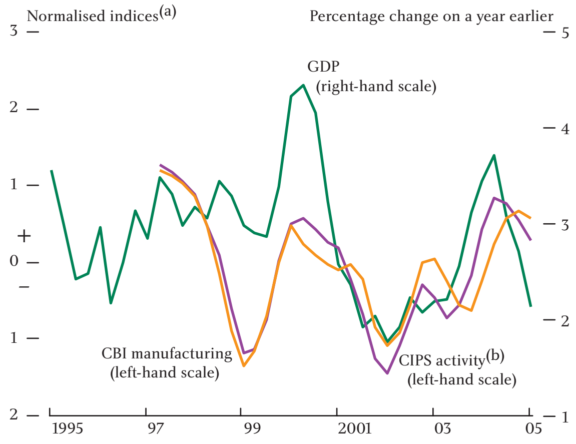


Chart 10
GDP and surveys of output



(a) Average index over preceding four quarters. Normalised by subtracting mean and dividing by standard deviation since 1996.
(b) Weighted sum of manufacturing output and services activity.

factors behind these GDP growth revisions are discussed in the box on page 340.

The initial underrecording of output growth undoubtedly led directly to errors in forecasting four-quarter GDP growth at short forecast horizons. For example, if we repeat regression (1) in Table D above, but add a variable that measures the degree to which the GDP data available at the time were misleading, then there is no longer any evidence of bias. The regression equation suggests that more than three quarters of the variance of the forecast error can be explained by data mismeasurement (Appendix C).

The key to understanding this result is to recognise that a one quarter ahead forecast of four-quarter GDP growth takes the first three quarterly growth rates as given by the data, and just requires a forecast of the final quarter. Any subsequent revision to the data will therefore feed one for one into the observed forecast error. So an assessment of short horizon forecasts should attempt to take account of revisions to the underlying data.

Could the tendency for initial GDP data to be revised have affected forecast errors over longer time horizons? It seems unlikely that two year ahead projections would be significantly affected. Forecasts of growth two years ahead are driven by factors like expected fiscal and monetary policy, external demand, and the expected evolution of asset prices, rather than current GDP growth. But forecast errors over shorter horizons are likely to have been affected.

Table H reports the correlation between forecast errors at different horizons, and the gap between the current estimate of GDP growth at the start of the projection and the estimate available at the time. All correlations are positive, suggesting that if actual growth was underrecorded, subsequent forecasts tended to be too weak. As expected, the correlation is strongest for one step ahead projections, and the correlation tends to fall as the horizon extends.

Table H
Correlation between errors forecasting GDP growth and revisions(a)

Forecast horizon	Correlation	Sample size
1	0.80	31
2	0.57	30
3	0.41	29
4	0.30	28
5	0.20	27

(a) To four-quarter growth at the start of the projection.

We turn now to the errors in 2002. MPC projections for growth in that year tended to be around 2.5% — somewhat higher than actual growth, which was around 2% (Chart 8). There was a pronounced slowing in growth from 2000. A major cause of the lower growth in GDP was the weakening of external demand. For example, a measure of world trade weighted according to UK export shares slowed from 12.4% growth in 2000, to just 3.1% in 2002. That led to a deceleration in UK exports, and a large negative net trade contribution (Table I).

Table I
Contributions to year-on-year GDP(a) growth

	2000	2002	Change
Household consumption	2.9	2.3	-0.7
Business investment	0.5	0.0	-0.4
Government consumption	0.7	0.9	0.1
Net trade	-0.1	-1.2	-1.1
GDP	4.0	2.0	-2.0
Memo: World trade(b)	12.4	3.1	-9.3

(a) Chained volume measures.
(b) Percentage changes on a year earlier. Weighted according to UK export shares.

The decline in world trade growth was sharper than the Committee had been expecting. For example, in the February 2001 *Inflation Report* the MPC noted that 'Growth in UK-weighted export markets is most likely to slow from around $10\frac{1}{2}$% in 2000, to just under 7% in 2001 and to 6% in 2002'. Given that around a third of UK output is exported, that would have been sufficient to explain the lion's share of the weaker-than-expected GDP growth, especially if one takes account of the likely second-round effects on UK business investment and household consumption of weaker exports.

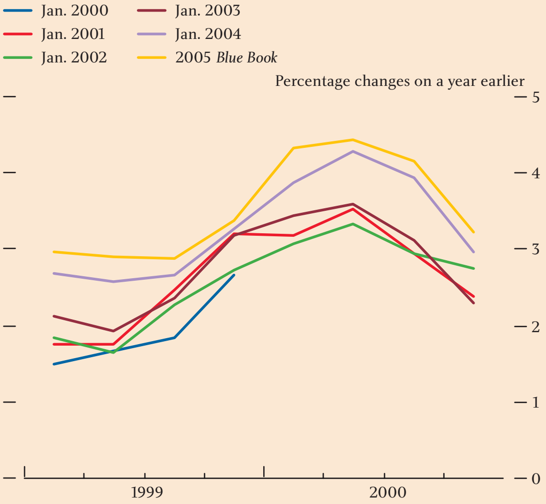
Revisions to GDP growth in 1999/2000

We have seen that the initial understatement of GDP growth in 1999 and 2000 appears to have led the MPC to under-forecast GDP growth. What lay behind the weakness of the initial ONS estimates?

The ONS publishes a first estimate of quarterly GDP growth within a month of the quarter finishing. Such timely first estimates are useful, but they are inevitably subject to revisions. They are based on incomplete information, and as more information comes in, the ONS is able to produce more accurate estimates. Furthermore, as the ONS builds a more complete picture of the economy, it will improve the way that GDP is measured, for example by using more accurate weights to sum components. The MPC is well aware of the likelihood of data revisions and takes account of that when making forecasts.(1) But as is clear from Chart 9, the revisions to the initial estimates of growth in 1999 and 2000 were particularly large.

From Chart A, we see that estimates of growth in 1999 and 2000 were revised several times as the ONS received more information. Growth was generally revised up from the previous estimates, apart from in January 2002. By far the largest upward revisions were made in 2003, several years after the MPC's projections for growth in 1999 and 2000 had been made.

Chart A
GDP: different vintages of ONS data



The 2003 revisions were made in the September release of the National Accounts when annual chain-linking was first introduced. The changes are discussed in a box in the November 2003 *Inflation Report*. The upward revisions to growth in 1999 and 2000 reflected new estimates of deflators of investment, exports and imports, as the ONS updated the weights it used. The changes were not directly related to annual chain-linking. So even though the MPC knew that the ONS was moving to annual chain-linking, these specific revisions were not predictable.

(1) See, for example, Ashley, Driver, Hayes and Jeffrey (2005).

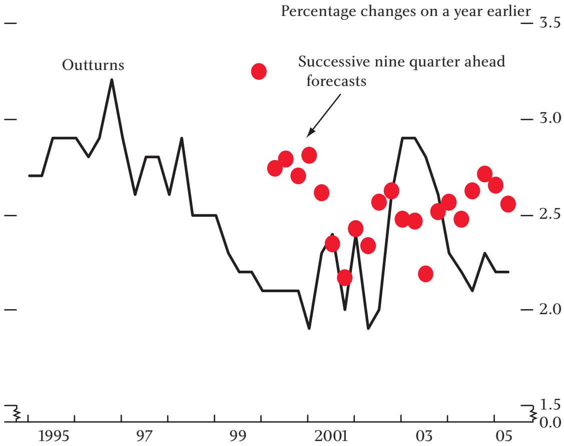
RPIX mean projections

Similar analysis of RPIX outturns and forecasts, as in Chart 11, suggests we should look at three episodes:

- the unexpectedly low inflation between 1999 and 2002;
- the unexpectedly high inflation in 2003; and
- the unexpectedly low inflation in late 2004 and early 2005.

Table J shows the contribution to RPIX inflation from four broad categories of goods and services. Changes in underlying inflation are monetary phenomena which reflect, among other things, expectations of monetary policy. But shocks in certain sectors may also affect aggregate inflation over short time horizons. The decomposition in Table J gives an indication of the type of shocks that may have had a temporary effect on inflation, allowing us to explore

Chart 11
RPIX inflation outturns and two year ahead mean forecasts(a)



(a) Mean projection based on market interest rates.

Table J
Contributions to RPIX annual inflation

	Average 1995–98	Average 1999–2002	2003	2004	2005 H1
Tradable goods[(a)]	0.8	-0.2	-0.1	-0.2	-0.5
Petrol, fuel and light	0.2	0.2	0.2	0.4	0.6
Housing[(b)]	0.5	0.8	1.2	1.0	1.1
Other (services and food)	1.3	1.4	1.5	1.0	0.9
RPIX	2.8	2.2	2.8	2.2	2.2

(a) Household and leisure goods, clothing and footwear, motor vehicles, alcohol and tobacco.
(b) Housing depreciation, rent and council taxes.

whether forecast errors have been attributable to those shocks.

The reduction in inflation between the periods 1995–98 and 1999–2002 is most noticeable in the category of 'tradable' goods. Indeed, price inflation for all other broad categories picked up (Table J). Is it possible that tradable goods prices were weaker than the MPC anticipated? The MPC did not make a forecast of the decomposition of RPIX prices. But, as is clear from Table K, the MPC's projections between 1997 and 1999 were based on assumptions for the exchange rate that turned out to be too low. For some projections, that gap was large. As a result, import prices tended to be lower than the Committee had been expecting. And this suggests that the sustained weakness in tradable goods price inflation, particularly between 1999 and 2000, probably led to some of the overprediction of inflation.

Table K
Exchange rate[(a)] projections and outturns

Forecast period[(b)]	Start point	Assumed level in two years[(c)]	Outturn	Percentage difference
1997	105.1	90.0	103.8	15.3
1998	104.7	99.7	106.4	6.7
1999	103.1	96.6	106.1	9.8
2000	106.1	104.6	105.7	1.1
2001	106.7	104.7	99.2	-5.3
2002	105.7	103.5	104.8	1.3
2003	98.6	96.7	102.5[(d)]	6.0

(a) As measured by the old IMF-based sterling effective exchange rate index, see Lynch and Whitaker (2004)
(b) Taken from August *Inflation Report* of each year.
(c) The projection in 1997 was based on judgement. In 1998 and 1999 the exchange rate was assumed to follow the path determined by UIP. From November 1999, it was assumed to follow the average between a flat path, and the UIP path.
(d) Average for July and August 2005.

In 2003, RPIX inflation was around 0.5 percentage points higher than the MPC had anticipated in earlier years (Chart 11). In that period, the housing components of the RPIX, and particularly housing depreciation (which is estimated from house price inflation) picked up sharply. House price inflation rose to over 20%, enough to raise RPIX inflation by 0.4 percentage points (Table G). These developments in the housing market were not anticipated by the MPC when preparing earlier projections. For example, in the November 2001 *Inflation Report*, 'house price inflation [was] projected to ease to a little below the growth rate of nominal earnings in the medium term'.[(1)] So the unexpected increases in RPIX inflation in 2003 largely reflected unexpectedly high house price inflation.

This exercise is less helpful in determining why RPIX inflation in 2004 and early 2005 was lower than expected by the MPC. As is clear from Table K, the exchange rate turned out to be considerably stronger than the assumptions in forecasts made in 2003. The appreciation of the exchange rate in early 2004 probably contributed to the higher rate of decline of imported goods, and that may explain part of the unexpectedly low RPIX inflation. But on the other hand, the MPC were not expecting such large increases in house prices or in oil, both of which pushed up on RPIX inflation in 2004 and 2005, relative to expectations. GDP growth was weaker than expected in early 2003, but stronger than expected in late 2003 and 2004. So the unexpectedly low inflation cannot be explained by unexpectedly weak demand. That raises the question, did potential supply grow faster than the MPC had expected?

The short-run trade-off between GDP growth and inflation

One factor that is common to MPC projections, and to external forecasts, is that between 1997 and 2003, RPIX inflation tended to be a little lower than expected, while GDP growth tended to be unexpectedly strong (Table D). To some extent that can be explained by unexpectedly weak import prices discussed above, which would tend to push down on inflation without necessarily depressing GDP growth. But there are several other candidate explanations relating to improvements in the UK supply side. Those include: developments in the UK retailing sector, which may have reduced the impact of demand on retail prices; government reforms to the labour market, which may have lowered the equilibrium rate of unemployment; and increased inward migration.

Since its inception in May 1997, as a result of monitoring the UK economy, and comparing outturns with mean projections, the MPC has made several adjustments to its assumptions about the relationship between activity and inflation. This continues to be an area of considerable uncertainty: see, for example, the discussion of risks on page 43 of the November 2004 *Inflation Report*. But further analysis of forecast errors should shed light on the accuracy of the MPC's assumptions on the supply side.

(1) Page 47.

Conclusions: an assessment of the MPC's projections

Fan charts

There is inevitably uncertainty around the outlook for the economy, and to communicate this, the MPC publishes its projections as fan charts. The purpose of this article has been to assess those fan chart projections.

We have considered various issues that arise when analysing fan chart projections. Perhaps the key point is that with only six years' worth of projections that can be compared with outturns, the sample is probably too small to draw firm conclusions. Furthermore, any assessment of the fan charts must take account of the likely serial correlation in forecast performance, which is a natural consequence of overlapping forecasts — ones that are repeated each quarter, and look many quarters into the future. Failure to take account of these and other issues runs the risk of drawing incorrect inferences about forecast performance.

With that in mind, we draw the following tentative conclusions:

- In general, between 1998 and 2005, GDP and RPIX outturns were dispersed broadly in line with the MPC's fan charts. So the fan charts gave a reasonably accurate summary of the risks and probabilities faced by the MPC.

- But for near-term projections of GDP growth, outturns were more dispersed than implied by the fan charts. To a large part that reflected unexpectedly big revisions to output data. In August 2005, the GDP fan charts were widened at short horizons to indicate better the level of uncertainty.

- For two year ahead projections of both GDP growth and inflation, outturns were less widely dispersed than implied by the MPC fan charts. But this is not strong evidence that those charts fanned out too widely. Each fan chart includes probabilities of many risks, and some of those risks will not occur. That does not mean that the risks were absent.

- The method used by the MPC to calibrate risk has tended to make fan charts narrower over time, in response to the recent economic stability. It is not clear that there is a strong case for making the latest inflation fan charts any narrower than they already are.

Mean projections

The probability of outturns following the mean projection from an MPC fan chart is small. But the profile of the mean is a key feature that summarises the shape of a fan chart, so it is instructive to assess how close outturns were to the mean projections. We draw the following conclusions:

- In the past five years, there have been periods when GDP and RPIX outturns were above or below the MPC's mean projections for several quarters in a row. But that it is not evidence of poor forecasting. For repeated forecasts that look many quarters into the future, it is what we might expect, even if the forecaster is making best use of all available information.

- In general, we conclude that the performance of the MPC's mean projections has been reasonably good when tested against a number of criteria, and when compared with other forecasters.

- Nevertheless, it is useful to ask what may have caused outturns to differ from the MPC's mean projections. The tendency for RPIX outturns between 1999 and 2002 to be lower than expected was related to the unexpected strength of the sterling exchange rate. And unexpectedly high inflation in 2003 reflected stronger-than-expected house price inflation. To some extent, the MPC's forecasts of GDP growth between 1998 and 2000 were too weak because the data available at the time, including both ONS data and business surveys, understated activity. And a principal cause of the unexpectedly weak GDP growth in 2002 was a sharper-than-expected slowdown in world demand.

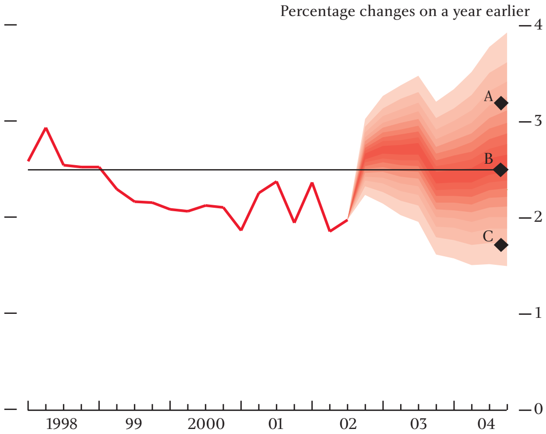
- The combination of unexpectedly low price inflation and unexpectedly strong output growth in the late 1990s and earlier this decade is consistent with various explanations. To a degree it appears to have reflected unexpectedly low import prices. But, in part as a result of monitoring outturns relative to its projections, the MPC also judges that the UK potential supply may have grown more rapidly than previously anticipated.

Appendix A: Density tests of the fan charts

This appendix sets out more formally how the tests of the fan charts are constructed. The fan chart describes the expected probability density function of inflation and GDP growth. Whatever the nature or shape of those probability distributions, we know that if they accurately describe reality, then over a large enough sample, outturns should be uniformly distributed across all probability bands. Our tests assess whether that has been true. For example, if the fan charts say that there is a 10% chance of RPIX inflation lying between 2.0% and 2.2%, then we examine whether 10% of outturns have been between 2.0% and 2.2%.

The key to testing density forecasts is to carry out a so-called 'probability integral transform' (PIT) of each outturn to capture the relationship between outturn and forecast. The PIT is the probability implied by the fan chart that an outturn would be equal or less than what was actually observed. For example, in Chart A, the PIT for outturn A would be 0.75, it would be 0.5 for outturn B, and 0.1 for outturn C. So the PIT identifies where the outturn fell relative to the fan chart bands. We refer to this measure as the percentile in which the outturn fell. If the fan charts give a good guide to the eventual dispersion of outturns, there should be an equal chance that an outturn falls in any percentile of the fan chart. Therefore a sample of the PITs of realised inflation or GDP growth should be uniformly distributed. This is demonstrated formally by Diebold, Gunther and Tay (1998).

Chart A
Different outturns relative to the fan chart



We employ two tests: the Kolmogorov-Smirnov (KS) test, and a likelihood ratio test suggested by Berkowitz (2001). The KS test is the standard test in the literature for comparing two distributions, by converting outturns so that they can be compared with the uniform distribution. The KS test focuses on the largest difference between the empirical distribution and the assumed distribution and tests how significant that difference is. Consider the example in Table A1. Outturns have been compared with the expected probability distribution, using the probability integral transform. The probabilities are listed by rank in the 'outturns row'. These are compared with the uniform distribution probabilities (row 2). The largest discrepancy from the uniform distribution is the first outturn, where the difference between probabilities is equal to -0.15.

Table A1
Example of applying KS test

Rank	1	2	3	4	5
Outturns	0.35	0.4	0.6	0.9	0.95
Uniform distribution	0.2	0.4	0.6	0.8	1.0
Difference	-0.15	0	0	-0.1	0.05

Kolmogorov and Smirnov derived a formula for testing the significance of this difference (See eg Kendall and Stuart (1979)). Let d = the difference, and let N = the number of observations. Then the p-value is given by:

$$Q\left(\left[\sqrt{N} + 0.12 + 0.11/\sqrt{N}\right]d\right)$$

Where $Q(x) = 2\sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2x^2}$

In the numerical example, feeding in a probability difference of 0.15, with a sample size of 5 gives a p-value of 0.9995. So in the example, there is no evidence to reject the null hypothesis that outturns were drawn from the expected distribution. The higher the probability number, the greater the confidence we have in rejecting the alternative hypothesis that the distribution of outturns was significantly different from that expected.

One problem with the KS test is that it makes no allowance for the possibility of serial correlation in forecast errors. If serial correlation causes a bias in the average forecast error, this will be reflected in the PITs of the outturns. The observed distribution will depart from the predicted one, which may cause the test to reject. And as we have discussed, with frequent forecasts that look a long way into the future, forecast errors are likely to be serially correlated.

The second test we apply is derived from Berkowitz (2001). Like the KS test, this is a test of equality of distributions. However Berkowitz suggested converting the sample of PITs of the outturns, using the inverse of the standard normal cumulative density function, so that they can be compared with the normal distribution rather than the uniform distribution. The advantage of moving to the normal distribution is that it is easy to construct standard likelihood ratio tests, which tend to be powerful. Berkowitz (2001) fits the following AR(1) model to the transformed probability values, which we refer to here as u_t:

$$u_t = \mu + \rho u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2) \qquad \textbf{(1)}$$

Two likelihood ratio tests are suggested. The first is a joint test of the equality of the observed and assumed (standard normal) distributions, and of independence of the observations. The second attempts to disentangle these hypotheses, by testing for independence of the observed u_t while allowing the mean and variance of the AR(1) process to differ from zero and unity respectively.(1)

For brevity, however, we do not report results for these tests. Instead we design a third test, in a similar vein, which focuses on the question most of interest here: allowing for the possibility of serial correlation in the relationship between fan charts and outcomes, have the fan charts given a good guide to subsequent outcomes? Because we allow for serial correlation to be present, we might expect inference from this test to be more reliable than that from the KS test. Specifically, we fit the AR(1) model to the observed u_t, and test the null hypothesis that the mean and variance of the u_t are zero and unity respectively, while allowing the autocorrelation coefficient to be freely estimated. The test is given by:

$$LR = -2\left[L\left(0, \hat{\rho}, \left(1-\hat{\rho}^2\right)\right) - L\left(\hat{\mu}, \hat{\rho}, \hat{\sigma}^2\right)\right] \sim \chi^2 \qquad \textbf{(2)}$$

Where $L(\mu, \rho, \sigma^2)$ is the standard log-likelihood function for an AR(1) process. We note that the distribution of the test statistic given in **(2)** is asymptotically valid, but is likely to be biased in small samples.

We allow for serial correlation in the form of an AR(1) as a simple approximation of the actual process we might expect. Forecast errors from a rational forecaster could be serially correlated as an MA process. For example, if quarterly growth rates of the variable being forecast are not serially correlated, rational forecasts of four-quarter growth rates should generate forecast errors that follow an MA(3) process. It is less clear what the precise autocorrelation structure of the PITs of the data should be, and so how good the AR(1) approximation is in practice. In this instance an AR(1) appears to give a reasonable approximation. Nevertheless, the results must be treated with caution.

(1) For more details of these tests see Berkowitz (2001).

Appendix B: How the size of the sample can affect statistical inference

This appendix reports a simple experiment to demonstrate how having a small sample can influence the probability of apparently finding significant evidence of forecast bias, even when it is not there.

Consider a series of forecast errors Y_t, which are serially correlated according to an MA(3) process:

$$Y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} \qquad \varepsilon \sim N(0,\sigma^2)$$

As discussed in the main text, it is likely that multi step ahead forecasts will generate serially correlated errors. More precisely, a rational forecaster should generate errors that follow an MA process. As the MPC's forecasts are of four-quarter growth rates, we would expect forecast errors to be at least MA(3) with unit roots. The longer the forecast horizon, and the more serially correlated the variable being forecast is, the more lagged terms we would expect in that MA process.

Because there is no constant in the equation, and each error term is mean zero, the errors are unbiased by construction.

A regression test for bias might involve regressing the errors on a constant:

$$Y_t = \alpha + u_t$$

and then testing to see if α is significantly different from zero. A naive test, that takes no account of serial correlation, might use a standard t test. Another alternative would be to use a t test, but based on standard errors calculated from a Newey-West variance-covariance matrix, which allows for serially correlated residuals.

Because we know the true distribution of the errors, we can use so-called 'Monte Carlo' techniques to assess the effectiveness of these tests. Using a random number generator we can generate a very large sample of data, and run repeated regression tests, to see how frequently they reject the hypothesis that the errors are unbiased.

We set the significance level of the tests at 5%, so if the tests were well specified they should only find significant evidence of bias 5% of the time.

From Table B1, it is clear that a combination of a small sample and serial correlation reduces the reliability of this simple test. Although by construction the implicit forecasts are not biased, the tests find evidence of bias too frequently. For the MA(3) process, standard t tests find evidence of bias around one third of the time, and when the sample is small, the Newey West adjustment does little to improve accuracy.

Table B1
The percentage of regression tests that found significant evidence of bias[(a)]

Sample size		Serial correlation			
		MA(3)	AR(1)[(b)]		
			$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 0.9$
20	t tests	37	11	25	69
	Newey West	24	9	18	57
1,000	t tests	33	8	25	67
	Newey West	10	6	6	14

(a) Each experiment involved running 1,000 regressions.
(b) $Y_t = \lambda Y_{t-1} + \varepsilon_t$.

The precise serial correlation of forecast errors cannot be known, so we also report the results for an AR(1) process with different values of lambda to illustrate the sensitivity of the test to the degree of serial correlation. Broadly speaking, a value of lambda between 0.5 and 0.8 should give a reasonable approximation of the degree of serial correlation we might expect for rational forecasts of four-quarter growth rates, depending upon the forecast horizon and the variable being forecast. As the results indicate, for such high degrees of serial correlation, the tests are not very reliable.

So when assessing the MPC's forecasts, it is important to bear in mind that the combination of a small sample of forecasts, and the likelihood of serially correlated errors, raises the probability of apparently finding significant evidence of bias when it is not actually present. The danger of misinterpreting the MPC's forecast record is increased further if analysis is confined to calculating simple average errors, or looking at charts.

Appendix C: Taking account of data revisions when testing for bias in GDP projections

In this appendix, we repeat the regression test reported in Table D in the main text, on one quarter ahead projections of four-quarter GDP growth. But we include a measure of the degree to which the initial GDP estimates available to the MPC, of quarterly growth in the preceding three quarters, have subsequently been revised. Because growth in the preceding three quarters enters the four-quarter growth calculation, there is likely to be a large direct effect from data revisions.

In Table C1 below we estimate the following equations:

$$Y_{t+1} - Y_t^{t+1} = \alpha \qquad \textbf{(1)}$$

$$Y_{t+1} - Y_t^{t+1} = \alpha + \beta R_t \qquad \textbf{(2)}$$

Where R_t is the revision to three-quarter GDP growth in the three quarters between t – 3 and t.

Table C1
Regression tests for bias on one quarter ahead projections of annual GDP growth(a)

	Constant (α)	Data revision (β)	R^2
Equation **(1)**	0.6 (0.00)	Not included	0.0
Equation **(2)**	0.0 (0.78)	1.2 (0.00)	0.8

(a) For mean production based on market expectations for interest rates, published between February 1998 and February 2005. The table cell is shaded orange if the estimated parameter is not significantly different from zero at 5% confidence levels and p-values are reported in brackets.

The results suggest that there has been a strong correlation between the MPC's forecast errors and the degree to which the initial data was subsequently revised. Once the test takes account of the extent to which the initial data were misleading, there is no longer significant evidence of bias. Indeed over three quarters of the variance of one quarter ahead forecast errors can be explained by the pattern of data revisions.

References

Allen, W and Mills, T (2005), 'How forecasts evolve — the growth forecast of the Federal Reserve and the Bank of England', *The National Institute Economic Review*, No. 193, pages 53–59.

Artis, M (1996), 'How accurate are the IMF's short term forecasts?', *IMF Working Paper no. 96/89*.

Ashley, J, Driver, R, Hayes, S and Jeffrey, C (2005), 'Dealing with data uncertainty', *Bank of England Quarterly Bulletin*, Spring, pages 23–29.

Bakhshi, H, Kapetanios, G and Yates, T (2003), 'Rational expectations and fixed-event forecasts: an application to UK inflation', *Bank of England Working Paper no. 176*.

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.

Benati, L (2005), 'The inflation-targeting framework from an historical perspective', *Bank of England Quarterly Bulletin*, Summer, pages 160–68.

Berkowitz, J (2001), 'Testing density forecasts with applications to risk management', *Journal of Business and Economic Statistics*, pages 465–74.

Britton, E, Fisher, P and Whitley, J (1998), 'The *Inflation Report* projections: understanding the fan chart', *Bank of England Quarterly Bulletin*, February, pages 30–37.

Clements, M (2004), 'Evaluating Bank of England density forecast of inflation', *The Economic Journal*, pages 844–66.

Clements, M and Hendry, D (2001), 'An historical perspective on forecast errors', *The National Institute Economic Review*, No. 177, pages 100–12.

Cogley, T, Morozov, S and Sargent, T (2004), 'Bayesian fan charts for UK inflation: forecasting and sources of uncertainty in an evolving monetary system', forthcoming, *Journal of Economic Dynamics and Control*.

Diebold, F X, Gunther, T A and Tay, A S (1998), 'Evaluating density forecasts with application to financial risk management', *Journal of Forecasting*, pages 285–95.

Diebold, F and Lopez, J (1996), 'Forecast evaluation and combination', in Maddala, G and Rai, C (eds), *The handbook of statistics*.

House of Lords Select Committee on Economic Affairs (2004), 'Monetary and fiscal policy: present successes and future problems'.

Kendall, M and Stuart, A (1979), *The advanced theory of statistics, volume 2: inference and relationship*, Charles Griffin & Co. Limited, London.

King, M (2003), 'Speech given at East Midlands development agency', *Bank of England Quarterly Bulletin*, Winter, pages 476–78.

King, M (2005), 'Remarks to the Central Bank Governors' panel', Jackson Hole conference 2005, available on the Bank of England website at www.bankofengland.co.uk/publications/speeches/2005/speech253.pdf.

Lynch, B and Whitaker, S (2004), 'The new sterling ERI', *Bank of England Quarterly Bulletin*, Winter, pages 429–41.

Melliss, C and Whittaker, R (1998), 'The Treasury forecasting record: some new results', *The National Institute Economic Review*, No. 164, pages 65–79.

Mitchell, J and Hall, S (2005), 'Evaluating, comparing and combining density forecasts using KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation', *National Institute Discussion Paper no. 253*.

Nordhaus, W (1987), 'Forecasting efficiency: concepts and applications', *The Review of Economics and Statistics*, Vol. 69, No. 4, pages 667–74.

Pagan, A (2003), 'Report on modelling and forecasting at the Bank of England', *Bank of England Quarterly Bulletin*, Spring, pages 60–88.

Poulizac, D, Weale, M and Young, G (1996), 'The performance of National Institute economic forecasts', *The National Institute Economic Review*, No. 156, pages 55–62.

Wallis, K (2004), 'An assessment of Bank of England and National Institute inflation forecast uncertainties', *The National Institute Economic Review*, No. 189, pages 64–71.