Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 24 (2008) 354–367
international journal of forecasting
www.elsevier.com/locate/ijforecast

Evaluating a three-dimensional panel of point forecasts: The Bank of England Survey of External Forecasters

Gianna Boero, Jeremy Smith, Kenneth F. Wallis*

Department of Economics, University of Warwick, Coventry CV4 7AL, UK

Abstract

This article provides a first analysis of the forecasts of inflation and GDP growth obtained from the Bank of England's Survey of External Forecasters, considering both the survey average forecasts published in the quarterly *Inflation Report*, and the individual survey responses, recently made available by the Bank. These comprise a conventional incomplete panel dataset, with an additional dimension arising from the collection of forecasts at several horizons; both point forecasts and density forecasts are collected. The inflation forecasts show good performance in tests of unbiasedness and efficiency, albeit over a relatively calm period for the UK economy, and there is considerable individual heterogeneity. For GDP growth, inaccurate real-time data and their subsequent revisions are seen to cause serious difficulties for forecast construction and evaluation, although the forecasts are again unbiased. There is evidence that some forecasters have asymmetric loss functions.

JEL classification: C53; E37

Keywords: Forecast surveys; Point forecasts; Density forecasts; Forecast efficiency; Asymmetric loss functions

1. Introduction

Every quarter since 1996, the Bank of England has asked a group of external forecasters for their views on some key macroeconomic indicators. Initially only forecasts of the official targeted measure of inflation were requested; subsequently, questions about GDP growth, the Bank's official interest rate, and the sterling effective exchange rate index were added. In general there are three questions about each variable, relating to three different forecast horizons. Of particular interest is the collection of expectations of future inflation and GDP growth, not only as point forecasts but also in the form of subjective probability distributions — so-called density forecasts. The Survey of External Forecasters (henceforth SEF) provides useful information on expectations outside the Bank about future economic developments and the likely achievement of the Bank's inflation target, which is supplied to the Bank's Monetary Policy Committee at its quarterly forecast meetings. The Committee meets monthly, and every three months prepares a forecast and the accompanying *Inflation*

* Corresponding author.
E-mail address: K.F.Wallis@warwick.ac.uk (K.F. Wallis).

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2008.04.003

Report. The quarterly *Inflation Report* also presents aggregate results from the current SEF.

The Bank of England has recently made available the individual SEF responses, suitably anonymised, for research purposes. This new source of survey data is comparable to the well known US Survey of Professional Forecasters, hitherto the only available source of time series of density forecasts in macroeconomics. The context is slightly different, since in the UK case the survey proprietor is itself a producer of published forecasts, to which its survey makes a useful input, as described below. Nevertheless, the new data offer opportunities to replicate and extend some of the studies undertaken on the US data. Our previous articles have derived measures of uncertainty and disagreement from the SEF data (Boero, Smith, & Wallis, in press), and analysed the revision process in repeated "fixed-event" forecasts of the same outcome (Boero, Smith, & Wallis, 2008). The present article provides a first analysis of the quality of the individual and aggregate point forecasts of inflation and GDP growth.

The data are described as a three-dimensional panel, following Davies and Lahiri (1995), because in addition to the usual two dimensions of a panel data set — individuals and time periods — we have multiple forecasts, reported in successive quarterly surveys, of particular inflation and GDP growth outcomes. A further dimension arises from the implicit availability of two point forecasts, one the reported point forecast, the other implicit in a measure of the location of the reported density forecast. We consider possible sources of differences between these two forecasts, and analyse the performance of both forecasts.

The remainder of this article is organised as follows. The structure of the survey and its use by the Monetary Policy Committee, together with the properties of the dataset and its dual point forecasts, are described in Section 2. The econometric framework used for forecast evaluation is presented in Section 3, and the empirical results follow in the next four sections. Section 4 deals with tests of unbiasedness and summary measures of forecast performance, Section 5 explores the possibility of asymmetry in the forecasters' loss functions, Section 6 presents evidence of individual heterogeneity, and Section 7 considers tests of forecast efficiency. Section 8 concludes.

2. The structure and use of the Survey of External Forecasters

2.1. Survey design

The institutions covered in the survey include City firms, academic institutions and private consultancies, and are predominantly based in London. The sample changes from time to time as old respondents leave or new survey members are included, and not every institution provides a forecast to the Bank every quarter, so the panel is incomplete. Although the survey members are individually anonymous, it is reasonable to assume that, collectively, there is considerable overlap with the professional forecasters regularly covered by services such as *Consensus Economics* and HM Treasury's monthly compilation, *Forecasts for the UK Economy*.

For the first two years the survey questions related only to inflation, defined with respect to the Retail Prices Index excluding mortgage interest payments (RPIX), in terms of which the official inflation target was defined; the survey definition switched to the Consumer Prices Index (CPI) from the February 2004 survey, following the change in the Bank's official targeted measure in December 2003. We date the surveys according to the date of the *Inflation Report* in which the aggregate results were published — February, May, August, and November — although the surveys were completed towards the end of the preceding month. Questions about point and density forecasts of GDP growth have appeared since February 1998, and about point forecasts of the official interest rate and the sterling exchange rate index since November 1999: these last two variables are not included in the present study, in the absence of density forecasts. The inflation section of a recent questionnaire is shown in Fig. 1; the GDP growth questions have the same format. Whereas the US Survey of Professional Forecasters offers respondents a choice of level or growth rates in reporting their point forecasts, which hinders the joint interpretation of point and density forecasts, here there is no ambiguity. Also, the point forecast question is immediately adjacent to the density forecasts, unlike the US survey, and the use of the neutral "central projection" term makes no attempt to specify which particular measure of central tendency of the density

YOUR COMPANY NAME (please complete): ________________________________

PROBABILITY DISTRIBUTION OF 12-MONTH CPI INFLATION OVER THE MEDIUM TERM

Please indicate the percentage probabilities you would attach to the various possible outcomes in 2005 Q4, 2006 Q4 and 2007 Q2. The probabilities of these alternative forecasts should of course add up to 100, as indicated.

PROBABILITY OF 12-MONTH CPI INFLATION FALLING IN THE FOLLOWING RANGES			
	2005 Q4	2006 Q4	2007 Q2
<1.0%			
1.0% to 1.5%			
1.5% to 2.0%			
2.0% to 2.5%			
2.5% to 3.0%			
≥ 3.0%			
TOTAL	100%	100%	100%

CENTRAL PROJECTION FOR 12-MONTH CPI INFLATION		
2005 Q4	2006 Q4	2007 Q2

Fig. 1. Bank of England questionnaire, May 2005 survey, inflation question.

forecast might be related to the respondent's point forecast.

Each quarterly survey since February 1998 asks for forecasts at three future points in time, as in the example in Fig. 1: the fourth quarter (Q4) of the current year; the fourth quarter of the following year; and the corresponding quarter two years ahead. (In the early "inflation-only" surveys, only the first two questions appeared.) This structure eventually delivers nine successive forecasts of a given Q4 outcome, which form a sequence of "fixed-event" forecasts, with the date of the forecast preceding the date of the outcome by 8, 7,…, 1, 0 quarters. Given that the survey goes out early in the quarter, when no data on current-quarter inflation and GDP growth are available, we treat these as h-step-ahead forecasts with the horizon h equal to 9, 8,…, 2, 1 quarters successively. In the more conventional time-series framework of constant-horizon forecasts, the third question delivers a quarterly series of nine-quarter-ahead forecasts, but the first two questions give only one observation per year at intermediate horizons, $h=4$ and 8 in February, $h=3$ and 7 in May, and so on. This focus on end-year targets is clearly more familiar to forecasters, since there are usually a few respondents who answer the first two questions but not the third question. Despite this, in May 2006 all three questions were switched to a constant-horizon format, focusing on the corresponding quarter one, two and three years ahead.

2.2. The Monetary Policy Committee's inflation forecasts

In an inflation targeting central bank, inflation forecasts are central to the conduct of policy, because it takes time for interest rate changes to affect inflation. "Inflation targeting implies inflation *forecast* targeting" (Svensson, 1997, p.1113; emphasis in original). In the United Kingdom in May 1997, the newly elected Labour Government granted operational independence to the Bank of England, through the newly created Monetary Policy Committee (MPC), to set interest rates in pursuit of an inflation target set by the Government. The MPC also assumed responsibility for the inflation forecasts previously produced by the Bank. Bean and Jenkinson (2001) describe the internal processes adopted by the MPC, covering the monthly policy round, the quarterly forecast round, and the preparation of the accompanying *Inflation Report*. In particular, at quarterly "draft forecast" meetings a few days before the associated MPC policy meeting, to help the MPC's overall assessment of its own forecast, a draft forecast is compared with external forecasts (Bean & Jenkinson, 2001, pp.439–440). These include the SEF, completed a few days earlier. Summary information from the survey, comprising average point forecasts and density forecasts, and the distribution of individual point forecasts, is then published in the *Inflation Report*, which features the MPC's density forecasts of inflation and GDP growth in the form of the famous fan charts.

An annual examination of the MPC's forecasting record has appeared in a box in the *Inflation Report* each August since 1999. For the point forecasts of inflation (density forecast means), these give an overall impression of performance that is regarded as satisfactory, with forecast errors averaging close to zero and a mean absolute error in one-year-ahead forecasts of 0.3 percentage points. No comparative evaluation has been reported to date, although such a study has recently been undertaken by Groen, Kapetanios and Price (2008), who compare the MPC's forecasts to a variety of inflation forecasting models, including linear and non-linear univariate models, and three- and five-variable VARs. This is described as a "real-time" evaluation, because the estimated forecasting models are conditioned on the dataset that was available at the time the MPC's forecasts were prepared. The results, for horizons $h=1$, 4 and 8, are striking: in no case does a model outperform the MPC forecasts, in terms of out-of-sample forecast RMSE. Groen et al. attribute the relatively poor performance of the statistical inflation forecasts to the importance of the judgment exercised by the MPC. In the present context this includes judgment of the weight to be given to external forecasts, and we note the result in Section 4.4 below that the SEF average point forecast of inflation in turn outperforms the MPC's forecast, so perhaps judgment could have been further improved. Casillas-Olvera and Bessler (2006) compare the published SEF average density forecasts two years ahead with those of the MPC, and find that the SEF does a better job than the MPC in terms of the Brier score for the inflation forecasts.

2.3. The dataset of individual SEF responses

The dataset of individual SEF responses made available by the Bank covers 39 surveys, beginning with the May 1996 survey and continuing to November 2005. Each respondent has an identification number, so that their individual responses can be tracked over time and their point and density forecasts can be matched. The total number of respondents appearing in the dataset is 48, one of whom stayed for only two years, while only one is present in all 39 surveys. To avoid complications caused by long gaps in the data, and to maintain the degrees of freedom at a reasonable level, most of our analyses that refer to individual forecasters are conducted on a subsample of 19 "regular respondents". These are respondents who each provided more than 70% of the total possible responses to the inflation and GDP growth questions over the available surveys, which number 39 for inflation and 32 for GDP growth.

2.4. Comparing point forecasts and density forecast means

The SEF density forecasts are reported as histograms, with respondents supplying their probabilities that future inflation or GDP growth will fall in each of a number of pre-assigned intervals, or bins. The histograms in the SEF data have rather few bins, between four and six, with the first and last being open-ended. A given density forecast implies a point forecast, as a measure of the location of the distribution, and we consider the mean, estimated by applying the standard formula, assuming that the reported probabilities are concentrated at the mid-points of the respective intervals, and that the open-ended intervals have an assumed finite width, equal to twice the width of the interior intervals. We first describe the nature of the differences between the reported point forecasts and the density forecast means, and then discuss the implications for the subsequent forecast evaluation exercises.

Summary evidence is presented in Tables 1 and 2, which report the percentage of cases in which the point forecast deviates from the density forecast mean by more than 0.2 percentage points, first for each regular respondent, then aggregated over this subsample, and finally for the full SEF sample. The two tables deal

Table 1
Divergences between point forecasts and density forecast means: inflation

	Question 1		Question 2		Question 3	
Individual	Above	Below	Above	Below	Above	Below
1	5.1	20.5	43.6	17.9	37.5	15.6
2	10.5	10.5	7.9	23.7	19.4	6.5
3	13.5	24.3	18.9	27.0	3.3	23.3
4	8.3	13.9	11.1	27.8	10.3	17.2
5	2.8	8.3	16.7	5.6	11.5	0
6	11.1	11.1	19.4	5.6	17.6	11.8
7	8.6	5.7	2.9	14.3	3.6	17.9
8	11.4	5.7	8.6	20.0	0	3.4
9	2.9	26.5	23.5	23.5	11.5	30.8
10	0	20.6	2.9	14.7	12.5	0
11	0	3.0	0	3.0	0	0
12	0	3.0	3.0	3.0	3.8	0
13	40.6	9.4	12.5	6.3	6.7	6.7
14	6.3	0	15.6	3.1	12.5	3.1
15	6.7	3.3	0	13.3	0	0
16	0	20.0	3.3	30.0	4.0	16.0
17	6.9	20.7	17.2	24.1	7.7	26.9
18	3.4	0	0	13.8	4.5	0
19	0	3.7	0	0	4.5	4.5
Subsample	8.2	12.4	12.1	15.9	10.5	12.2
Full sample	6.2	11.5	10.2	14.1	9.4	9.8

Percentage of cases (across all available time periods) in which the point forecast lies above (below) the density forecast mean by more than 0.2 percentage points.

respectively with the inflation and GDP growth forecasts. The full-sample information is also given by Boero et al. (in press, Table 1), who compare the results with those of Engelberg, Manski and Williams (in press). These authors document the extent to which reported point forecasts in the US Survey of Professional Forecasters deviate from calculated measures of the location of the reported density forecasts, and observe that in such cases, forecasters are inclined to present "favourable" scenarios, in the sense that, more often than not, their point forecasts anticipate lower inflation and higher output growth than is indicated by the measures of the location of their density forecasts. For the GDP growth forecasts, this tendency increases as the forecast horizon increases, which is in accordance with the finding in the literature on subjective probability judgments that optimism increases with the forecast horizon (see Milburn, 1978, for an early discussion and experimental evidence).

The full-sample SEF data in Tables 1 and 2 show a greater tendency towards similarly "favourable" scenarios in respect of GDP growth than inflation, with this again increasing as the forecast horizon (or question number) increases. However, in the two-year-ahead inflation forecasts the deviations are more evenly balanced, suggesting that, in an inflation targeting regime, a favourable scenario is one in which the official target is achieved in the medium term, and this may lead to a positive or negative adjustment of an initial forecast. The aggregate data for the regular respondent subsample show the same patterns, although these data mask considerable variation across the 19 individual rows of each table. We note that the interpretation of Engelberg et al. (in press) that "forecasters who skew their point predictions tend to present rosy scenarios" implicitly uses the density forecast as the base for the comparison, whereas the SPF questionnaire first asks for point forecasts of several variables and then, on a separate

Table 2
Divergences between point forecasts and density forecast means: GDP growth

	Question 1		Question 2		Question 3	
Individual	Above	Below	Above	Below	Above	Below
1	40.6	6.3	40.6	15.6	34.4	3.1
2	29.0	9.7	54.8	12.9	50.0	0
3	60.0	16.7	76.7	6.7	76.7	0
4	17.2	13.8	31.0	10.3	67.9	0
5	13.8	6.9	24.1	3.4	23.1	11.5
6	31.0	13.8	55.2	0	68.8	12.5
7	39.3	10.7	57.1	7.1	48.1	3.7
8	25.9	3.7	50.0	7.7	45.5	0
9	29.6	33.3	29.6	22.2	23.1	26.9
10	21.4	25.0	39.3	21.4	42.9	7.1
11	11.5	0	7.7	0	14.3	0
12	15.4	7.7	11.5	0	12.0	0
13	36.7	6.7	73.3	3.3	63.3	6.7
14	18.8	9.4	25.0	0	15.6	18.8
15	23.1	11.5	57.7	3.8	62.5	0
16	33.3	8.3	54.2	8.3	32.0	12.0
17	20.0	4.0	32.0	8.0	11.5	11.5
18	18.2	0	68.2	0	63.6	0
19	4.5	4.5	9.1	0	22.7	0
Subsample	27.9	10.5	44.8	8.0	43.0	6.7
Full sample	27.4	8.8	44.2	6.6	46.6	4.6

Percentage of cases (across all available time periods) in which the point forecast lies above (below) the density forecast mean by more than 0.2 percentage points.

page, for density forecasts of inflation and GDP growth. The SEF question shown in Fig. 1 gives less priority to one forecast or the other, and differences in forecasters' loss functions provide an alternative interpretation of the individual divergences summarised in Tables 1 and 2.

The optimal point forecast is the mean of the density forecast under the conventional quadratic loss function, but is different from the mean under other symmetric loss functions if the density is asymmetric. For a symmetric linear loss function the optimal point forecast is the median, and for a bounded "all or nothing" loss function the optimal point forecast is the mode of the density forecast. And if the loss function is asymmetric, the optimal point forecast again differs from the density forecast mean.

Regression tests of the divergences between the reported point forecast and the density forecast mean find a highly significant negative coefficient on the skewness of the forecast density, both in the pooled subsample and in many individual cases, for both inflation and GDP growth. This holds in regressions with the point forecast as the dependent variable, including with the density forecast mean as a regressor, or with the divergence between them as the dependent variable, or in an ordered probit regression based on the classification of the divergence used in Tables 1 and 2. Recalling that positive skewness implies mean>median>mode, this result is consistent with forecasters reporting a point forecast closer to one that is optimal under some loss function other than the standard quadratic loss function, such as in the two examples given above. The individual cases in which a significant negative skewness coefficient does not appear are those in which relatively few divergences greater than ±0.2 are observed, including individuals with several zero entries in Tables 1 and 2. These differences across individuals are suggestive of differences in their loss functions. This heterogeneity merits further investigation in its own right, and also suggests that forecast evaluations should be conducted not only in the standard framework, which is founded on squared error loss, but also in the framework proposed by Elliott, Komunjer and Timmermann (2005), which allows more general loss functions. They find that allowing for asymmetric loss can significantly change the outcome of empirical tests of forecast rationality, and some of their procedures are included below.

3. The econometric framework

We adopt the notational convention of the three-dimensional panel introduced by Davies and Lahiri (1995), and denote the h-step-ahead forecast of the outcome at time t made by individual i as F_{ith}. Thus, the data are sorted first by individual, $i=1,..,N$, with $N=19$ in our regular respondents subsample, then by target period $t=1,\ldots,T$, and finally by horizon, with the earliest, longest-horizon forecast listed first. If t refers to the first, second or third quarter of a year then there is only one forecast, with horizon $h=H=9$, whereas if t refers to a fourth quarter then we have nine forecasts, with $h=H,H-1,\ldots,1$ (except in the initial years). With these definitions and the time series structure described in Section 2.1, the forecast F_{ith} is based on an information set dated $t-h$, denoted I_{t-h}, and the forecast is elicited by the survey carried out in quarter $t-h+1$.

With the actual outcome for the variable of interest denoted A_t, Davies and Lahiri (1995) decompose the forecast error $A-F$ as

$$A_t - F_{ith} = \phi_i + \lambda_{th} + \varepsilon_{ith}. \tag{1}$$

The second component λ_{th} is common to all individual forecast errors — it has no i subscript — and it represents the cumulative effect on A_t of uncorrelated period-by-period aggregate shocks since the forecast was made. The first and third components of the forecast error are specific to individual forecasters, separating a possible systematic effect or individual bias ϕ_i from an idiosyncratic non-autocorrelated error, ε_{ith}, which might reflect individual sentiment or "animal spirits", and mishandling of or inadequacies in the initial information set. The extension by Davies (2006) relaxes the assumption that the individual forecaster's bias ϕ_i is independent of the forecast horizon h, so the decomposition becomes

$$A_t - F_{ith} = \phi_{ih} + \lambda_{th} + \varepsilon_{ith}. \tag{2}$$

General expressions for estimates of the forecast error components in Eq. (1) are

$$\hat{\phi}_i = \frac{1}{TH}\sum_{t=1}^{T}\sum_{h=1}^{H}(A_t - F_{ith}), \tag{3}$$

$$\hat{\lambda}_{th} = \frac{1}{N}\sum_{i=1}^{N}\left(A_t - F_{ith} - \hat{\phi}_i\right), \tag{4}$$

$$\hat{\varepsilon}_{ith} = A_t - F_{ith} - \hat{\phi}_i - \hat{\lambda}_{th}. \tag{5}$$

In the extended decomposition of Eq. (2), estimates of the horizon-specific individual biases are

$$\hat{\phi}_{ih} = \frac{1}{T}\sum_{t=1}^{T}(A_t - F_{ith}), \quad h = 1, \ldots, H. \tag{6}$$

If the number of point forecast errors was the same for all horizons, as is implicit in the above definition, then Eq. (3) would give

$$\hat{\phi}_i = \frac{1}{H}\sum_{h=1}^{H}\hat{\phi}_{ih}.$$

This does not hold in the SEF dataset by virtue of the structure of the questions, irrespective of additional complications caused by missing observations. To represent the SEF structure described above, we distinguish between the quarters, $q=1,\ldots,4$, by writing the time index as $t=4(y-1)+q$, where $y=1,\ldots,Y$ indicates the year, so that the specific form of Eq. (3) is

$$\hat{\phi}_i = \frac{1}{Y}\sum_{y=1}^{Y}\Bigg(\frac{1}{3}\sum_{q=1}^{3}\left(A_{4(y-1)+q} - F_{i,4(y-1)+q,9}\right) + \frac{1}{9}\sum_{h=1}^{9}\left(A_{4y} - F_{i,4y,h}\right)\Bigg), \tag{3'}$$

with a corresponding specialisation of Eq. (6).

Tests of the unbiasedness and efficiency of forecasts under squared error loss can be based on regressions of observed forecast errors on various regressors. The simplest is the test of the null hypothesis of zero mean forecast error, or unbiasedness of individual forecasts, and the estimate $\hat{\phi}_i$ in Eq. (3) is equal to the coefficient in a regression of the forecast error on an intercept term. Tests of efficiency or rationality check for the absence of a correlation between the forecast error and the information available at time $t-h$ by regressing the error on candidate variables from this information set, which may include previous forecast errors.

The regression context is convenient for considering questions of inference, and in particular the calculation of the relevant forecast error covariance matrix, denoted $\mathbf{\Sigma}$, and associated regression coefficient covariance matrices, given as $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ in the usual least squares regression notation. Davies and Lahiri (1995, Section 2.2, 2.3) consider the covariance between two typical forecast errors

$$\begin{aligned}&\mathrm{cov}(A_{t_1} - F_{i_1t_1h_1}, A_{t_2} - F_{i_2t_2h_2})\\ &\quad = \mathrm{cov}(\lambda_{t_1h_1} + \varepsilon_{i_1t_1h_1}, \lambda_{t_2h_2} + \varepsilon_{i_2t_2h_2}).\end{aligned}$$

They develop expressions for the $NTH \times NTH$ covariance matrix $\mathbf{\Sigma}$ and its estimation, with an extension (Davies & Lahiri,1995, Section 5) to accommodate observed heteroskedasticity. These can readily be adapted to the structure of the SEF described above. Missing data are handled by appropriately compressing the data and covariance matrices (Davies & Lahiri, 1995, p.213).

4. Unbiasedness tests and forecast RMSEs

4.1. Point forecasts

Analyses of each regular respondent's point forecast errors are summarised in Tables 3 and 4. Table 3 contains results for inflation, using the appropriate definition of the target variable, namely the Retail Prices Index excluding mortgage interest payments (RPIX) for the surveys up to November 2003, and the Consumer Prices Index (CPI) from the February 2004 survey. These series are never revised after first publication. Table 4 contains results for GDP growth, first using real-time data as the actual outcome data to define the forecast errors, and then replacing these with revised ("historical") data, as of the August 2007 issue of *Economic Trends*. In all cases the outcome data extend to 2005Q4, so the summary statistics in the two panels of Table 4 include the contribution of very recent forecast

Table 3
Analysis of inflation point forecast errors

Individual	RMSE	$\hat{\sigma}^2_{\varepsilon_i}$	$\hat{\phi}_i$	SE	t-ratio
1	0.570	0.244	−0.047	0.120	−0.390
2	0.400	0.065	−0.013	0.112	−0.118
3	0.359	0.044	−0.001	0.110	−0.013
4	0.386	0.107	0.026	0.113	0.226
5	0.416	0.069	−0.101	0.111	−0.908
6	0.314	0.048	−0.074	0.111	−0.665
7	0.424	0.071	−0.221	0.111	−1.991**
8	0.302	0.020	−0.086	0.123	−0.698
9	0.350	0.074	0.024	0.121	0.198
10	0.529	0.243	0.043	0.120	0.360
11	0.409	0.086	0.034	0.111	0.310
12	0.535	0.081	−0.269	0.111	−2.428**
13	0.343	0.029	0.056	0.113	0.495
14	0.407	0.061	0.053	0.116	0.459
15	0.328	0.029	−0.021	0.106	−0.195
16	0.302	0.029	−0.030	0.116	−0.256
17	0.358	0.063	−0.017	0.106	−0.165
18	0.461	0.082	−0.026	0.123	−0.216
19	0.331	0.040	−0.002	0.112	−0.022

Note to Tables 3–6: ***indicates significance at the 1% level, **indicates significance at the 5% level, and *indicates significance at the 10% level.

errors which are defined with respect to outcome data which are, as yet, relatively similar. The maximum possible number of inflation forecasts is 98, supplied by only one forecaster, as noted above, and the maximum possible number of GDP growth forecasts is 84.

Columns 2–6 of the tables report the root mean squared forecast errors (RMSE), the variance $\hat{\sigma}^2_{\varepsilon_i}$ of the idiosyncratic errors estimated as in Eq. (5), the individual bias $\hat{\phi}_i$ given in Eq. (3), its standard error calculated from the appropriate covariance matrix, and the resulting t-ratio. The significance of the t-test of the null hypothesis of zero individual bias at the 10, 5 or 1% levels is indicated by one, two or three asterisks respectively.

4.1.1. Inflation

The mean forecast error is negative for 13 of the 19 regular respondents, indicating a general tendency to overpredict inflation. The mean error is significantly different from zero at the 5% level in two cases, both being cases of a significant negative bias. The larger of these, and the largest absolute bias overall, is approximately a quarter of one percentage point. Although this is statistically significant, in practical terms it is small, compared to the average outcome of 2.5% for RPIX inflation (which is exactly equal to the target value) and the deviation of ±1 percentage point from the target that triggers the requirement of a public explanation by the Governor of the Bank of England. (The range of the quarterly observations of the annual RPIX inflation rate is 1.9–3.2%.)

4.1.2. GDP growth

The upper panel of Table 4 is based on GDP growth forecast errors defined with reference to real-time data on

Table 4
Analysis of GDP growth point forecast errors

Individual	RMSE	$\hat{\sigma}^2_{\varepsilon_i}$	$\hat{\phi}_i$	SE	t-ratio
Errors calculated from real-time GDP data					
1	0.678	0.260	−0.112	0.229	−0.490
2	0.587	0.064	−0.083	0.222	−0.376
3	0.715	0.101	−0.196	0.223	−0.877
4	0.721	0.164	0.062	0.223	0.276
5	0.948	0.309	−0.154	0.217	−0.707
6	0.562	0.201	−0.188	0.223	−0.844
7	0.872	0.159	−0.097	0.224	−0.434
8	0.730	0.099	−0.225	0.224	−1.003
9	0.796	0.391	−0.129	0.233	−0.555
10	0.841	0.343	0.168	0.233	0.720
11	0.668	0.114	−0.289	0.221	−1.309
12	0.873	0.191	−0.275	0.221	−1.246
13	0.671	0.100	−0.239	0.234	−1.020
14	0.747	0.132	−0.153	0.228	−0.673
15	0.708	0.111	−0.036	0.213	−0.170
16	0.723	0.090	−0.054	0.242	−0.224
17	0.749	0.151	−0.072	0.216	−0.334
18	0.708	0.147	−0.082	0.238	−0.342
19	0.848	0.185	−0.010	0.222	−0.044
Errors calculated from revised GDP data					
1	0.964	0.287	0.326	0.335	0.971
2	0.925	0.070	0.353	0.329	1.072
3	0.959	0.108	0.235	0.332	0.710
4	1.019	0.150	0.506	0.328	1.542
5	1.192	0.252	0.302	0.314	0.960
6	0.913	0.117	0.254	0.329	0.773
7	1.133	0.170	0.368	0.330	1.114
8	0.993	0.117	0.213	0.332	0.639
9	0.899	0.354	0.287	0.337	0.852
10	1.150	0.367	0.561	0.339	1.653*
11	0.898	0.124	0.133	0.324	0.412
12	1.055	0.179	0.202	0.322	0.629
13	0.837	0.145	0.106	0.348	0.304
14	1.074	0.145	0.302	0.338	0.892
15	1.043	0.118	0.433	0.309	1.402
16	1.056	0.104	0.416	0.360	1.158
17	1.071	0.171	0.333	0.313	1.067
18	1.081	0.155	0.437	0.349	1.253
19	1.151	0.187	0.444	0.324	1.371

the actual outcomes. These show a more pronounced tendency towards overprediction than the inflation forecast errors, with negative mean forecast errors for all but two of the regular respondents, but no mean error is significantly different from zero. However, the picture changes when the evaluation is based on revised GDP data, as is shown in the lower panel of the table. The general effect of revisions to the national accounts over this period has been to increase the preliminary estimates of GDP growth, and the magnitude of the revisions is sufficient to turn all the negative mean errors in the upper panel into positive mean errors in the lower panel. Although none of these is significantly different from zero at the 5% level, the increase in the mean error of individual 10 from the small positive number in the upper panel is sufficient to make it significantly different from zero at the 10% level in the lower panel. As noted above, revisions to recent data are not yet complete; nevertheless, over the full-sample period, the average upward revision in quarterly observations on the annual GDP growth rate is 0.57 percentage points. (The range of the observations is 0.75–3.7% in the real-time data and 1.6–4.4% in the revised data.) The relative difficulty of tracking the revised data is indicated by the increase in absolute value of the mean errors from the upper to the lower panel, and the associated increase in RMSE.

The general difficulty of forecast construction and evaluation in the face of data revisions is well appreciated in the forecasting literature (for a recent survey see Croushore, 2006). The specific difficulties facing the Monetary Policy Committee with respect to recent revisions in UK GDP data are highlighted in two boxes in the August 2005 issue of the Bank of England *Inflation Report*. If it is thought that the revised data are closer to the truth, and that they should be the forecaster's objective, then the difficulty is in deciding where to start, since the current initial conditions will be subject to revision. This led the MPC to substantially widen the uncertainty bands around the current-quarter and next-quarter forecasts of GDP growth from August 2005, and to publish backcasts and nowcasts of GDP growth, with uncertainty bands, from August 2007.

4.2. Density forecast means

Corresponding analyses of the errors in the density forecast means give results similar to those presented for the reported point forecasts in Tables 3 and 4, and do not merit the inclusion of an additional pair of tables. For inflation, the mean forecast error is negative for all but two of the 19 individual forecasters, although only one of these is now larger than -0.2 and significant at the 5% level. There are increases in RMSE as often as decreases between the two forecasts, although there is a slight decrease in the overall level, mostly as a result of the two forecasters with the largest RMSE values in Table 3 — individuals 1 and 12 — showing rather smaller RMSEs of their density forecast means. For GDP growth, we observe the same effects as above when comparing evaluations against real-time data with those using revised data, namely an increase in mean errors such that they are all positive in the latter case, and associated increases in RMSE. Once more, however, no mean error is significantly different from zero at the 5% level.

4.3. Horizon-specific individual biases

Although the preceding analyses show no forecast biases that merit further investigation, the assumption that the individual forecaster's bias, ϕ_i in Eq. (1), is independent of the forecast horizon h may be unduly restrictive. We relax this assumption by considering the extended error decomposition in Eq. (2), and estimating horizon-specific individual biases. For each regular respondent we calculate $\hat{\phi}_{ih}$ as in Eq. (6), and consider tests of their individual significance and tests of their equality over h.

For the inflation point forecast errors, we reject the null hypothesis of equality of forecast bias over different horizons at the 5% level for four of our 19 regular respondents. These four all have two or three horizon-specific $\hat{\phi}_{ih}$ coefficients which are significantly different from zero at this level, although for three of them the overall bias shown in Table 3 is not significant; the fourth case is individual 7. Individual 12 is a counterexample, with a significant overall bias, resulting from several significant biases at different horizons, and non-rejection of the null hypothesis of equality across horizons.

The GDP growth real-time data forecast errors show two individuals whose insignificant overall bias masks variation by horizon that is significant at the 5% level. However, overall there are only two horizon-specific individual $\hat{\phi}_{ih}$ coefficients that are significantly different from zero. Substantially different results are obtained with the revised outcome data. Again there are only two individuals with significant variation across horizons, but 17 out of 19 individuals have two or three significant $\hat{\phi}_{ih}$ coefficients, all of which relate to short horizons. Once more these findings are masked by the absence of significant overall biases in Table 4, but they represent a further manifestation of the difficulties discussed above. The impact of data revisions on forecast performance is greatest at short horizons, because it is here that accurate initial conditions are of the greatest importance; at longer horizons forecasts tend to return to trend.

A similar pattern of results is found for the density forecast means, for both variables.

4.4. The SEF average and the MPC's forecasts

By way of a summary, akin to the study by Casillas-Olvera and Bessler (2006), Table 5 presents comparable results for the average forecasts across all survey respondents and the forecasts of the Monetary Policy Committee. Each quarterly *Inflation Report* presents the average point forecasts from the current Survey of External Forecasters and, with rather fuller discussion, the MPC's current forecasts. The latter covers all intermediate future quarters up to the forecast horizon, and we extract the fan chart modes that correspond to the three survey questions from the Bank's forecast spreadsheets. The mode, or most likely outcome, is the MPC's preferred central projection. It is seen that the forecasts are unbiased in all cases.

Table 5
Comparison of SEF average and MPC forecasts

	No. errors	RMSE	Mean	SE	*t*-ratio
Inflation					
SEF average	98	0.286	−0.057	0.110	−0.518
MPC	98	0.368	0.005	0.109	0.046
GDP growth (real-time data)					
SEF average	84	0.611	−0.079	0.222	−0.354
MPC	84	0.634	−0.139	0.225	−0.617
GDP growth (revised data)					
SEF average	84	0.947	0.359	0.333	1.077
MPC	84	0.911	0.299	0.334	0.895

Comparisons of RMSE for forecasts of GDP growth show little difference between the SEF average and MPC forecasts, whether real-time or revised data are used as the actual outcome data. Again, we have a clear indication of the increased difficulty of forecasting the revised GDP growth data, as discussed above. For the inflation forecasts, which receive greater attention in an inflation targeting context, the RMSE comparison clearly favours the SEF average forecast, as was noted in Section 2.2 above. It is also notable that the SEF average forecast RMSE is smaller than any individual regular respondent's RMSE shown in Table 3. Although the 19 regular respondents do not always enter the published survey average, which also includes other less regular respondents, this result supports the familiar advantage to be gained by forecast pooling. The same result does not hold for the GDP growth forecasts, Table 4 showing a few individual RMSEs that are smaller than that of the SEF average forecast in each case, again suggesting ambiguity regarding the forecasters' target measure.

5. Asymmetries in forecasters' loss functions

We return to the question raised at the end of Section 2 with the evidence of the point forecast errors analysed above, in the context of the generalised loss function proposed by Elliott et al. (2005), which is

$$L(p,\alpha) = [\alpha + (1-2\alpha) \times 1(A-F<0)]|A-F|^p \quad (7)$$

for integer p and $0 \le \alpha \le 1$. With $\alpha = 0.5$ this gives the familiar symmetric linear and quadratic loss functions for $p = 1,2$ respectively, while their asymmetric counterparts are obtained if $\alpha \ne 0.5$. For a given value of p, Elliott et al. develop estimators of α and tests of forecast rationality under asymmetric loss. When analysing mean forecast errors, an estimate $\hat{\alpha} \ne 0.5$ "can be interpreted as justifying biased forecasts by adjusting the loss function to make them optimal", as they say (2005, p.1113). Given the ambiguity in the definition of GDP growth forecast errors, with respect to real-time or revised data, we study only the inflation forecasts.

Estimates of α are presented in Table 6, first relaxing the symmetry assumption of the quadratic

Table 6
Asymmetry parameter estimates: inflation

Individual	Quadratic	Linear
1	0.56	0.50
2	0.52	0.49
3	0.50	0.50
4	0.46	0.43
5	0.65**	0.53
6	0.64**	0.54
7	0.82***	0.69***
8	0.67***	0.49
9	0.46	0.39**
10	0.45	0.35***
11	0.45	0.36**
12	0.82***	0.58
13	0.40	0.35***
14	0.42	0.39**
15	0.54	0.49
16	0.56	0.47
17	0.53	0.43
18	0.54	0.46
19	0.51	0.34***

loss function implicit in the foregoing analysis, then also moving to an asymmetric linear loss function. Estimated coefficients that lead to rejection of the null hypothesis that $\alpha=0.5$ are indicated by asterisks, as above. Relevant expressions for the standard error of the estimated coefficient are given by Elliott et al., and we implement them under the null for the purpose of constructing an appropriate *t*-statistic.

Maintaining a quadratic function, it is seen that there are five significant departures from symmetry, all with $\hat{\alpha}>0.5$. These individuals have the five largest absolute biases in Table 3, with the biases all being negative. A value of α in excess of 0.5 implies that positive forecast errors incur greater loss than equivalent negative forecast errors; that is, there is a greater fear of underprediction than of overprediction, and hence the tendency to overpredict demonstrated by these individuals may be loss-minimising. In the face of general underestimation of the amplitude of peaks and troughs found in the literature, these forecasters are anxious not to miss a peak in inflation, but are less concerned about missing a trough.

Under a linear specification, it is immediately noticeable that smaller values of $\hat{\alpha}$ are obtained. The reduction is such that, of the five individuals significantly above 0.5 in the quadratic case, only one remains significantly so under linearity. At the same time, six individuals have values of $\hat{\alpha}$ in the range 0.34–0.39, all significantly different from 0.5. It is difficult to relate these results to the calculations reported in Table 3, which rest on the standard squared error loss framework, and further investigation through other means seems warranted. The calculations reported in this section can be relatively sensitive to outliers, but the main case in which these are evident is that of individual 1, discussed below, who is not of major concern in this section.

6. Individual heterogeneity

The good performance of the SEF average forecast shown in Table 5 and the general unbiasedness of the regular respondents' individual forecasts mask considerable variation in their forecast performance. As is seen in Table 3, the RMSE of inflation point forecasts ranges from 0.30 to 0.57, while the corresponding ranges for GDP growth forecasts (Table 4) are 0.56 to 0.95 if forecast errors are defined with reference to real-time outcome data, and 0.84 to 1.19 if revised outcome data are used. On closer inspection, the maximum inflation RMSE of 0.57 for individual 1 is heavily influenced by a small number of early longer-horizon forecasts of inflation in excess of 4%, which turned out to be much too high. This pessimism occurred prior to the establishment of the policy of inflation targeting by an independent central bank in mid-1997 and the establishment of its credibility among these forecasters, as documented in our previous article (Boero et al., in press). This overestimation of inflation was not sustained long enough to cause a significant bias, but shows up in this individual's maximum idiosyncratic variance. The decomposition in Eq. (1) identifies three components of the forecast error, and hence of its root mean square, and we undertake more systematic comparisons across the two components that are specific to individual forecasters. Results are presented in turn for mean and variance effects, respectively individual biases and idiosyncratic error variances.

Individual heterogeneity with respect to mean forecast errors is assessed by testing the equality of $\hat{\phi}_1,\ldots,\hat{\phi}_{19}$, as shown in the relevant columns of Table 3 and the upper and lower panels of Table 4. The corresponding test statistics are 1.60, 0.55 and 0.32, respectively. The total number of forecasts entering the calculation is $n=1561$ for inflation and $n=1322$ for GDP growth, and the 5%

critical value of the $F(18, n)$ distribution is 1.61. Thus, the null hypothesis of equal mean forecast errors, or individual homogeneity in this respect, is not rejected, although the inflation forecast mean errors are very close to rejection at this level.

A test of individual heterogeneity with respect to variance is based on the variance of the idiosyncratic error ε_{ith} defined in Eq. (1): the null hypothesis is that the variances $\sigma^2_{\varepsilon_i}$ are equal across individuals. From the estimates reported in Table 3 and the two panels of Table 4, we obtain test statistics of 5.96, 5.05 and 4.46, respectively. A comparison with the 1% critical value of F (18, n) of 1.95 indicates highly significant individual heterogeneity with respect to forecast error variance for both variables.

Overwhelming rejections of individual homogeneity are reported by Davies and Lahiri (1999) for inflation forecasts in the US Survey of Professional Forecasters, the survey dataset that is most comparable to the SEF. Their individual analyses are based on the performance of 45 forecasters who responded more than 50% of the time, supplying forecasts one to four quarters ahead for up to 89 target dates (1969Q4–1991Q4). Individual biases are much more prevalent than in our sample, with 12 of the 45 forecasters having mean errors significantly different from zero at the 5% level (or higher), whether preliminary or revised outcome data are used, and whether one quarter horizon forecasts or all forecasts together are considered. Individual homogeneity with respect to idiosyncratic error variance is strongly rejected. However, their sample period involved a considerably different inflationary experience to ours, including as it does the "Great Inflation" of the 1970s, when there were well-documented forecast failures (McNees, 1979). The significant individual biases are all cases of positive mean errors, that is, underpredictions of inflation. In contrast, RPIX inflation during our sample period was in the range 1.9–3.3%, as noted above. Nevertheless, the common finding of individual heterogeneity with respect to idiosyncratic error variance indicates that some respondents in each survey sample are better at forecasting than others.

7. Tests of efficiency

The rational expectation or efficient forecast of the outcome A_t under squared error loss is its expected value, conditional on information available at time $t-h$, and hence the null hypothesis of forecast efficiency is written

$$F_{ith} = E(A_t|I_{t-h}),$$

assuming that I_{t-h} contains "all available" information. Standard tests of forecast efficiency then check the orthogonality of the forecast errors to variables which the researcher considers to be likely members of the forecast information set. If a variable is found which is correlated with the forecast error, and could thus have improved the forecast, then the rejection of the null hypothesis is conclusive. On the other hand, failure to reject efficiency may simply reflect the researcher's failure to find the information that the forecaster had overlooked. Since the information set includes past A, F and $A-F$, a simple test that is often used is to check for the absence of autocorrelation of order h or more in the forecast errors. Given the relative preponderance of forecasts at $h=9$ in the SEF dataset and the low power of high-order autocorrelation tests, together with the double loss of degrees of freedom for autocorrelation coefficients caused by missing observations, we do not pursue this possibility.

To test the efficiency of the inflation forecasts, we consider four variables that are likely members of the information set: the latest available inflation outcome at the time the survey was carried out, the most recent forecast by the MPC, the most recent SEF average forecast, and the real-time output gap or inflationary pressure measure of Garratt, Lee, Mise and Shields (2006). (We use the published average SEF forecast rather than the individual's own previous forecast because, again, missing observations cause a double loss of degrees of freedom.) We first test the admissibility of this instrument set in a multiple regression of the forecast on the four candidate variables. This gives significant regression results at the 5% level or (mostly) higher for 15 of the 19 regular respondents, from which we conclude that these variables are valid candidate variables for testing efficiency. Proceeding to a regression of the forecast errors on these variables, we find a significant regression result, and hence a rejection of efficiency, in three cases, namely individuals 1, 7 and 12, while 16 of the 19 regular respondents pass this test.

For the GDP growth forecasts, the longer publication delay in national accounts data implies that the latest observation of the outcome that can be included in I_{t-h} is A_{t-h-1}, not A_{t-h}. In these circumstances we also include the monthly GDP estimate produced by the National Institute of Economic and Social Research as a candidate variable (Mitchell, Smith, Weale, Wright & Salazar, 2005), together with the most recent MPC and SEF average forecasts, as in the case of inflation. The multiple regression of the forecast on these four candidate variables gives a significant regression result for 10 out of 19 individuals, which is less strong evidence in favour of these instruments than in the case of inflation, perhaps reflecting greater difficulties caused by data delays and the forecasters' treatment of potential revisions. Nevertheless, on turning to efficiency tests we find overwhelming evidence against the null hypothesis, with rejections for 16 individuals if errors are defined with respect to the real-time data, and for all but one individual if errors are defined with respect to the revised data. Among the candidate variables, the most recent MPC forecast has a strong positive correlation with many individual forecasts, and a strong negative correlation with the forecast error, leading to the rejection of efficiency; the MPC's own forecast fails this test. This suggests that among individual respondents there is, unfortunately, too great a tendency to "follow my leader" in this respect. This is not a feature of the inflation forecasts.

8. Conclusion

This article provides a first evaluation of the forecasts of inflation and GDP growth obtained from the Bank of England's Survey of External Forecasters, considering both the survey average forecasts published in the quarterly *Inflation Report*, and the individual survey responses recently made available by the Bank. The survey was initiated by the Bank in 1996 to provide independent input to its own inflation forecast processes, which became the responsibility of the Monetary Policy Committee following its establishment in 1997. The published SEF average forecasts of inflation are seen to have outperformed the MPC's forecasts: in this respect the pacemaker is winning the race.

Access to the individual survey responses allows replication of the point forecast evaluations performed on several similar survey datasets by several authors. A distinguishing feature of the SEF, however, is that it also collects subjective probability assessments or density forecasts, in which respect its only existing counterpart is the long-established Survey of Professional Forecasters in the United States. This allows a dual evaluation, of both the reported point forecasts and the alternative point forecasts implied by a measure of the location of the density forecasts. As with the US data, differences between the two forecasts in the SEF can be interpreted as a tendency towards reporting point forecasts that represent "favourable" or optimistic outcomes, although an alternative interpretation of forecasters' behaviour in terms of asymmetric loss functions is preferred.

In tests of unbiasedness of the inflation and GDP growth forecasts, both the survey average and a subsample of individual forecasters present an overall picture of good performance. As is often remarked in discussions of the performance of the Monetary Policy Committee over its first ten years, however, this was a relatively calm period for the UK economy, and presented no serious problems to forecasters, in the form of major turning points, for example. Behind the general picture lies considerable individual heterogeneity, shown not only by the failure of standard tests of equality of idiosyncratic error variances, but also by further evidence of different degrees of asymmetry in forecasters' loss functions. In the face of this individual heterogeneity, the good performance of the survey average forecasts of inflation is another example of the benefits of forecast pooling.

The inflation forecasts also perform well in tests of efficiency, whereas the GDP growth forecasts do not. This finding may be related to the familiar difficulty of measurement, with inaccurate real-time national accounts data and their subsequent extended revision process causing difficulties for forecast construction and evaluation. Data revisions substantially change the overall impression of the performance of forecasts of GDP growth at short horizons, as is also the experience of the Monetary Policy Committee.

Many questions remain to be explored in the context of the SEF dataset, replicating and extending an already large body of empirical literature on forecasting. Several of our findings prompt questions about the individual forecasters' methods and objectives, the exploration of which would be worthwhile.

Acknowledgements

We are grateful to Colin Ellis, Debbie Lockhart and Rebecca Capuano (Bank of England) for assembling and helping to clean the survey dataset. Readers wishing to gain access to the data should write to the Publications Editor, Inflation Report and Bulletin Division, Bank of England, Threadneedle Street, London EC2R 8AH, UK. The first version of this article was presented at the Royal Economic Society Annual Conference held at the University of Warwick on 11–13 April 2007; the helpful comments of anonymous referees and an Associate Editor are gratefully acknowledged.

References

Bean, C., & Jenkinson, N. (2001). The formulation of monetary policy at the Bank of England. *Bank of England Quarterly Bulletin*, *41*(4), 434−441.

Boero, G., Smith, J., & Wallis, K.F. (in press). Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *Economic Journal*.

Boero, G., Smith, J., & Wallis, K. F. (2008). Here is the news: Forecast revisions in the Bank of England Survey of External Forecasters. *National Institute Economic Review*, *203*, 68−77.

Casillas-Olvera, G., & Bessler, D. A. (2006). Probability forecasting and central bank accountability. *Journal of Policy Modeling*, *28*, 223−234.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 961–982). Amsterdam: North-Holland.

Davies, A. (2006). A framework for decomposing shocks and measuring volatilities derived from multi-dimensional panel data of survey forecasts. *International Journal of Forecasting*, *22*, 373−393.

Davies, A., & Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, *68*, 205−227.

Davies, A., & Lahiri, K. (1999). Re-examining the rational expectations hypothesis using panel data on multi-period forecasts. In C. Hsiao, M. H. Pesaran, K. Lahiri, & L. F. Lee (Eds.), *Analysis of Panels and Limited Dependent Variable Models* (pp. 226−254). Cambridge: Cambridge University Press.

Elliott, G., Komunjer, I., & Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, *72*, 1107−1125.

Engelberg, J., Manski, C.F., & Williams, J. (in press). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*.

Garratt, A., Lee, K.C., Mise, E., & Shields, K., (2006). *Real time representation of the UK output gap in the presence of uncertainty*. Working Papers in Economics & Finance 0618, Birkbeck College.

Groen, J.J.J., Kapetanios, G., & Price, S., (2008). *Real time evaluation of* Inflation Report *and* Greenbook *forecasts for inflation and growth*. Presented at the Royal Economic Society Annual Conference, University of Warwick, March 2008.

McNees, S. K. (1979). The forecasting record for the 1970s. *New England Economic Review*, 33−53, September/October 1979.

Milburn, M. A. (1978). Sources of bias in the prediction of future events. *Organizational Behavior and Human Performance*, *21*, 17−26.

Mitchell, J., Smith, R. J., Weale, M. R., Wright, S., & Salazar, E. L. (2005). An indicator of monthly GDP and an early estimate of quarterly GDP growth. *Economic Journal*, *115*, F108−F129.

Svensson, L. E. O. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European Economic Review*, *41*, 1111−1146.

Forecast sources cited

Bank of England *Inflation Report* (quarterly): http://www.bankofengland.co.uk/publications/inflationreport/index.htm

Consensus Economics: http://www.consensusforecasts.com/

HM Treasury *Forecasts for the UK Economy* (monthly): http://www.hm-treasury.gov.uk/economic_data_and_tools/data_index.cfm

US Survey of Professional Forecasters: http://www.philadelphiafed.org/econ/spf/index.cfm