Calibration and Resolution Diagnostics for Bank of England Density Forecasts

John W. Galbraith
Department of Economics
McGill University
855 Sherbrooke St. West
Montreal, QC, Canada
H3A 2T7

Simon van Norden
Affaires Internationales
HEC Montréal
3000 Ch. de la Côte Ste. Catherine
Montreal, QC, Canada
H3T 2A7

Version of 26 August 2008: preliminary and incomplete

Abstract

This paper applies new diagnostics to the Bank of England's pioneering density forecasts (*fan charts).* We compute their implicit probability forecast for annual rates of inflation and output growth that exceed a given threshold (in this case, the target inflation rate and 2.5% respectively.) Unlike earlier work on these forecasts, we measure both their calibration and their resolution, providing both formal tests and graphical interpretations of the results. Our results both reinforce earlier evidence on some of the defects of these forecasts and provide new evidence on their information content.

Key words: calibration, density forecast, probability forecast, resolution

* The *Fonds québécois de la recherche sur la société et la culture* (FQRSC), the Social Sciences and Humanities Research Council of Canada (SSHRC) and CIRANO (*Centre Interuniversitaire de recherche en analyse des organisations)* for support of this research.

1. Introduction

Since their introduction in the 1993 Inflation Report, the Bank of England's probability density forecasts ("fan charts") for inflation, and later output growth, have been studied by a number of authors. Wallis (2003) and Clements (2004) studied the inflation forecasts and concluded that while the current and next-quarter forecast seemed to fit well, the year-ahead forecasts significantly overestimated the probability of high inflation rates. Elder, Kapetanios, Taylor and Yates (2005) found similar results for the inflation forecasts, but also found significant evidence that the GDP forecasts do not accurately capture the true distribution of risks to output growth at very short horizons.[1] They also explored the role of GDP revisions in accounting for GDP forecast errors and noted that the dispersion associated with predicted GDP outcomes was increased as a result of their research. Dowd (2008) examined the GDP fan charts and found that while short-horizon forecasts appear to poorly capture the risks to output growth, results for longer horizon forecast are sensitive to the vintage of data used to evaluate the forecasts. In contrast with this apparent consensus, the August 2008 Inflation Report considered the performance of the fan charts over the previous six quarters and indicated that GDP growth has been improbably lower than would be expected from the fan charts, while inflation has been improbably higher.

Throughout this work, the focus has been on whether the risks suggested by the Bank's density forecasts are well matched (in a statistical sense) by the frequency of the various inflation and output growth outcomes. A key advantage of these forecasts is that, by conveying the full density function, they allow forecast users to compute the probability of an outcome lying in a particular range of interest. It is of course possible that the overall correctness of the conditional density specification may be rejected, but that the density forecasts nonetheless give good probability predictions for particular problems of interest, such as the probability of inflation lying in a given band, or falling above a threshold. In the present paper we work with the implied probabilistic forecasts for such questions of interest: we compute the implied probability forecasts of failing to achieve an inflation target, or for GDP growth falling below a given threshold. The methods that we use to do so can of course be applied to implied forecast probabilities for other questions of interest, simply by integrating under different regions of the density forecast.

Our approach to forecast density evaluation therefore differs in two respects from that in the previous literature. First, we simplify the forecasting problem to one with a binary outcome; we consider the probability that the outcome will be no greater than some given threshold level. This is similar in spirit to Clements' (2004) examination of the fan chart's implied interval forecasts.[2] Choices of different thresholds allow one

[1] Noting the hazards of drawing firm conclusions from small samples, these authors suggested that "the fan charts gave a reasonably good guide to the probabilities and risks facing the MPC [monetary policy committee]."

[2] The Bank of England also examines such interval forecasts from time to time. For

to focus on different parts of the forecast distribution and thereby gain additional insights into forecast performance for certain types of outcomes. With a binary outcome corresponding with a problem of interest we are able to investigate the the calibration of the forecasts, that is, the degree to which predicted probabilities correspond with true probabilities of the outcome. Second, we examine the ability of forecasts to discriminate between different outcomes, as opposed to producing similar probability forecasts in circumstances which are in fact different in relevant respects. The methods that we use have not been applied to these forecasts in previous evaluative work, and provide a very different lens through which the performance of the fan charts can be judged.

In contrast with earlier evaluations, our preliminary results provide strong evidence of a miscalibration of the inflation forecasts at very short horizons, even though the degree of miscalibration appears to be small. Despite the much shorter sample available for the GDP forecasts, we also find significant evidence of miscalibration, and its magnitude appears to be much larger than for inflation. Results on the discriminatory power of the forecasts shows that inflation forecasts appear to have important power up to horizons of about one year, while that of GDP forecasts is much less and is of little use beyond a one quarter horizon.

2. Predictive density evaluation

Let X be a random variable with realizations x_t and with probability density and cumulative distribution functions $f_X(x)$ and $F_X(x)$ respectively. Then for a given sample $\{x_t\}_{t=1}^T$, the corresponding sample of values of the CDF, $\{F_X(x_t)\}_{t=1}^T$, is a U(0,1) sequence. This well-known result (often termed the probability integral transform of $\{x_t\}_{t=1}^T$ is the basis of much predictive density testing, following pioneering work by Diebold, Gunther and Tay (1998).

These authors noted that if the predictive density $\hat{f}_X(x)$ is equal to the true density, then using the predictive density for the probability integral transform should produce the same result. This allows us to test whether a given sequence of forecast densities could be equal to the true sequence by checking whether $\{\hat{F}_X(x_t)\}_{t=1}^T$ (i.e. the sequence CDFs of the realized values using the forecast densities) is U(0,1). If this sequence is assumed to be independent, the U(0,1) condition is easily tested with standard tests (such as a Kolmogorov-Smirnov one-sample test.)

The independence is unrealistic in many economic applications, however. In particular, violation is almost certain for multiple-horizon forecasts as the $h-1$ period overlap in horizon-h forecasts induces an MA$(h-1)$ process in the forecast errors. The inferential problem is therefore more difficult: test statistic distributions are affected by the form of dependence. Nonetheless, it is often instructive to plot the histogram (or empirical CDF) of the sequence to check for economically important deviations from the U(0,1). We do so below.

---

example, see Table 1 (p. 47) of the August 2008 Inflation Report.

3. Calibration and resolution

The calibration and/or resolution of probabilistic economic forecasts have been investigated by a number of authors, including Diebold and Rudebusch (1989), Galbraith and van Norden (2007), and Lahiri and Wang (2007). The meteorological and statistical literatures contain many more examples; some recent contributions include Hamill et al. (2003) and Gneiting et al. (2007). Probability statements can of course readily be derived from densities; calibration addresses the question of whether probability statements have their stated meanings. Resolution measures information content of forecasts, in the sense of discriminatory power. These concepts are defined precisely below.

3.1 Methods and definitions

Following the notation of Murphy and Winkler (1987), let x be a 0/1 binary variable representing an outcome and let $\hat{p} \in [0, 1]$ be a probability forecast of that outcome. Forecasts and outcomes may both be seen as random variables, and therefore as having a joint distribution; see e.g. Murphy (1973), from which much subsequent work follows. Since the variance of the binary outcomes is fixed, it is useful to condition on the forecasts: in this case we can express the mean squared error $E((\hat{p} - x)^2)$ of the probabilistic forecast as follows:

$$E(\hat{p} - x)^2 = E(x - E(x))^2 + E_f(\hat{p} - E(x|\hat{p}))^2 - E_f(E(x|\hat{p}) - E(x))^2, \tag{3.1}$$

where $E_f(z) = \int z f(z) dz$ with $f(.)$ the marginal distribution of the forecasts, $\hat{p}$. Note that the first right-hand side term, the variance of the binary sequence of outcomes, is a fixed feature of the problem and does not depend on the forecasts. Hence all information in the MSE that does depend on the forecasts is contained in the second and third terms on the right-hand side of (3.1).

Numerous summary measures of probabilistic forecast performance have been suggested, including loss functions such as the Brier score (Brier, 1950) which is a MSE criterion, and measures such as the calibration, probability integral transform, and resolution. The latter are not intended to be loss functions; that is, forecasts may perform well on these measures while performing poorly in other respects and overall. Nonetheless such measures are often of interest because they are valuable in interpreting forecasts, and taken together may provide a good overall picture of forecast performance. The MSE is of course only one of many possible loss functions, and is inappropriate in many circumstances. However it is nonetheless interesting to relate it to concepts of independent interest, such as the calibration and resolution; each of these can be shown to emerge from a decomposition of the forecast MSE.

We will call the first of the terms in (3.1),

$$E_f(\hat{p} - E(x|\hat{p}))^2, \tag{3.2}$$

the (mean squared) *calibration error:* it measures the deviation from a perfect match between the predicted probability and the true probability of the event when a given

forecast is made.[3] If for any forecast value $\hat{p}_i$ the true probability that the event will occur is also $\hat{p}_i$, then the forecasts are perfectly calibrated. If for example we forecast that the probability of a recession beginning in the next quarter is 20%, and if over all occasions on which we would make this forecast the proportion in which a recession will begin is 20%, and if this match holds for all other possible predicted probabilities, then the forecasts are perfectly calibrated. Note that perfect calibration can be achieved by setting $\hat{p} = E(x)$, the unconditional probability of a recession, since the expectation is taken over the possible values or range of values that the probability forecast can take on.

Calibration has typically been investigated using histogram-type estimates of the conditional expectation, grouping probabilities into cells; Galbraith and van Norden (2008) use smooth conditional expectation functions estimated via kernel methods, and we use these methods (described in detail in that paper) below.

The last term on the right-hand side of (3.1), $E_f(E(x|\hat{p}) - E(x))^2$, is called the forecast *resolution*, and measures the ability of forecasts to distinguish among relatively high-probability and relatively low-probability cases. Note again that the expectation is taken with respect to the marginal distribution of the forecasts. If resolution is high, then in typical cases the conditional expectation of the outcome differs substantially from its unconditional mean: the forecasts are successfully identifying cases in which probability of the event is unusually high or low. The resolution enters negatively into the MSE decomposition; high resolution lowers MSE. To return to the previous example, the simple forecast that always predicts a 5% probability of recession, where 5% is the unconditional probability, will have zero resolution. Perfect forecasts would have resolution equal to variance (and zero calibration error, so that MSE = 0).

The calibration error has a minimum value of zero; its maximum value is 1, where forecasts and conditional expectations are perfectly opposed. The resolution also has a minimum value of zero, but its maximum value is equal to the variance of the binary outcome process. In order to report a more readily interpretable measure, scaled into $[0, 1]$, we divide the resolution by the variance of the binary outcome process. Let n be the number of observed forecasts and $\mu = E(x)$; then the maximum resolution achievable arises where there are $n\mu$ 1's and $n - n\mu$ 0's constituting the sequence $E(x|\hat{p})_i$. The resulting maximum total is $n\mu(1-\mu)^2 + n(1-\mu)\mu^2$. Divide by n for the mean; this quantity is then the maximum resolution and is also equal to the variance of a 0/1 random variable with mean μ. Therefore

$$\frac{E_f(E(x|\hat{p}) - \mu)^2}{\mu(1-\mu)^2 + (1-\mu)\mu^2} \in [0, 1]. \tag{3.3}$$

[3]This quantity is often called simply the 'calibration' or 'reliability' of the forecasts. We prefer the term calibration *error* to emphasize that this quantity measures deviations from the ideal forecast, and we will use 'calibration' to refer to the general property of conformity between predicted and true conditional probabilities.

The information in the resolution is correlated with that in the calibration; the decomposition just given is not an orthogonal one (see for example Yates and Curley 1985). However the resolution also has useful interpretive value which we will see below in considering the empirical results.

4. Data and forecasts

The Bank of England's Inflation Report provides probabilistic forecasts of inflation and, more recently, output growth in the form of 'fan charts,' indicating quantiles of the forecast distribution. Fan charts for RPIX inflation were published from 1993Q1 to 2004Q1, when they were replaced by CPI Inflation fan charts. The GDP fan chart was first published in the 1997Q3 report. In addition to providing forecast distributions for roughly 0 to 8 quarters into the future, from the 1998Q1 forecast onwards these are provided conditional on the assumption of both fixed interest rates, and a "market-expectation-based" interest rate profile. The two assumptions typically provide very similar results, but we will nonetheless present most results below for both sequences of density forecasts.

While these charts provide only a visual guide to the degree of uncertainty that the Bank of England (BoE) associate with their forecasts, they are based on an explicit parametric model of forecast uncertainty, as documented by Brittan, Fisher and Whitley (1998) and Wallis (2004), among others. Forecast errors are assumed to follow a 'two-piece normal' or 'bi-normal' distribution, whose behaviour is completely characterized by three parameters: a mean μ, a measure of dispersion σ, and a parameter which controls skewness, γ.[4] These parameters therefore allow us to estimate the implied forecast probabilities that inflation or GDP growth would exceed any given threshold level.[5]

The forecasts evaluated in this paper, therefore, are of both GDP and inflation, made given either fixed interest rates, or market interest rates. In any of these four cases, we derive (from the densities represented in fan charts) implied probabilities of the variable of interest lying on one side of these targets; for GDP forecasts, we use a threshold of 2.5% growth, and for inflation, we use the inflation target. We therefore translate the information in the fan chart densities into forecasts of the probability of

[4]See Wallis (2003, Box A on p. 66) for a discussion of the bi-normal distribution and its alternative parameterizations. Spreadsheets containing the parameter settings for all of the published fan charts are publicly available on the BoE's web site (presently at http://www.bankofengland.co.uk/publications/inflationreport/irprobab.htm).

[5]Like the normal distribution, the bi-normal lacks an exact closed-form formula for its cumulative distribution function (CDF). We therefore estimated the implied forecast probabilities by numerical integration from probability density function using the GAUSS function `intquad1`. The resulting CDF estimates appeared to be accurate to at least 0.001. Additional details and code are available from the authors upon request.

not exceeding the given threshold, and evaluate these probabilistic forecasts.

While a few sets of forecast densities are available at horizons exceeding eight quarters, the sample sizes involved are small. We therefore report results for nine horizons, zero (the 'nowcast' of the eventual current-quarter release) through eight.[6]

With the four cases just noted, we then have thirty-six sets of forecasts for evaluation.

In this preliminary version we measure inflation and output growth outcomes using the 2008Q2 vintage data series. In a future draft, we will explore the sensitivity of our results to data revision.

5. Empirical results

The empirical results of this paper are summarized primarily through graphical diagnostic, which we will now describe.

5.1 Descriptive statistics

Figures 1 and 2 plot the probabilistic outcomes (i.e. $\hat{F_X}(x_t)$ as defined above). Each point corresponds with a particular inflation or output growth outcome as forecast at the horizon given on the horizontal axis; the larger (green) dots represent cases in which the eventual outcome was below the relevant threshold, while the smaller (blue) dots are cases in which the outcome was above threshold.

Ideal forecasts would have assigned probabilities near one to all the large green dots and probabilities near zero to the smaller blue dots. Instead, for GDP growth we observe several high probability blue dots and low probability green dots (see horizons 2-4 in particular) which indicate "surprises." We also find most outcomes clustered in the center of the probability range at horizons 4-8, suggesting that the GDP growth forecasts lack resolution at these horizons. The probabilistic outcomes for inflation show similar features with respect to changes across horizons, but at the shorter horizons we see a more marked concentration of large green dots at the higher probabilities and small blue dots at the lower, suggesting that the inflation forecasts had more discriminatory power than the GDP forecasts, at least at the short horizons.

Figures 3 and 4 plot histograms of the probability integral transforms for each of the sets of forecasts; as these are U(0,1) under the null of correct specification of the conditional density, these histograms should show roughly the same proportion of observed forecasts in each of the ten cells. While some sampling variation is of course inevitable, these patterns are in general far from conformity with this condition. GDP growth forecasts (Figure 3) often show an excessive number of values in the highest cell (near 1) at short horizons, an insufficient number at long horizons, and an insufficient number of values near zero at virtually all horizons. Inflation forecasts (Figure 4) show better conformity with the desired pattern of uniformity, but some of the same tendency is observable. Note that an insufficient number of values of the probability

[6]The figures below also show some estimates based on longer horizons, although the sample sizes available are small.

integral transform near the extremes is an indication of forecast densities that are too dispersed: actual outcomes occur near the tail of the forecast density less often than would arise with the true conditional density, and therefore observed outcomes tend to be in intermediate regions of the relevant CDF.

In the following subsection we explore the calibration and resolution of the forecasts, to understand better the source of these apparent problems observed in the descriptive statistics.

5.2 Calibration and resolution

Figures 5 and 6 plot the estimated conditional expectation of outcome given forecast, which for correctly calibrated forecasts would lie along the line $E(x|\hat{p}) = \hat{p}$, i.e. the 45 degree line. Again, see Galbraith and van Norden (2008) for a description of the methods used to construct these estimates, including bandwidth choice.

Figure 5 shows these conditional expectations for the GDP growth forecasts at each forecast horizon. Deviations from perfect calibration (the 45-degree line) are widespread and often large. Most strikingly, *high* predicted probabilities of growth falling below threshold correspond to *low* observed frequencies of this outcome for virtually all horizons. Calibration is much more reasonable for probabilities in the 0-0.5 range, but falls dramatically beyond predicted probabilities of around 0.8. Calibration for the small sample of very long horizon forecasts looks disastrous, with a strongly negative slope everywhere. In contrast, calibration of the inflation forecasts is appears reasonable at all predicted probabilities, even for the small sample of longer-horizon forecasts available in the market interest rate case.

The results of formal tests of the null hypothesis of correct calibration (that is, $E(x|\hat{p}) = \hat{p}$) are given in Table 1. At short horizons, correct calibration is decisively rejected for both GDP growth and inflation forecasts. At longer horizons the results are mixed, with differences emerging between the fixed interest rate and market interest rate forecasts. Despite the apparently dramatic mis-calibration visible in the figures for GDP forecasts, we are often unable to reject the null of correct calibration for the market-interest-rate forecasts. This may be related to the relatively small number of sample points lying at high forecast probabilities, which would give the test low power to detect such deviation. Alternatively, it could be due to the highly non-linear evidence of miscalibration shown in the graphs. Our test is based on a linear alternative hypothesis and so may have reduced power against some non-linear alternatives. In contrast, the graphs for inflation showed that the forecast calibration was much more nearly linear, which should give our tests higher power. This may therefore explain the relative high number of rejections of the null despite the relatively better fit shown in the inflation graphs.

Figures 7 and 8 record graphical information about the resolution of the probability forecasts. For each horizon (only horizons up to four quarters are shown), each figure presents a pair of empirical CDF's: that of probability forecasts in cases for which the eventual outcome was below threshold, and in cases in which the eventual outcome was above threshold. In a near-ideal world, probability forecasts should be near one

in cases where the outcome turned out to be below threshold, and near zero when the outcome turned out to be above. In that case the two CDF's would lie close to the lower horizontal axis in the first case, and close to the upper horizontal axis in the second. More generally, good probability forecasts will discriminate effectively between the two possible outcomes, and the two empirical CDF's should be widely separated. At longer horizons, the value of conditioning information declines and this separation becomes more difficult to achieve; we therefore expect to see the pairs of CDF's less widely separated at longer horizons.

This pattern of reduced separation with horizon is in fact readily observable; at 3-4 quarter horizons, we observe little separation on either forecast series. However, at shorter horizons, we observe a clear distinction between the GDP growth and inflation forecasts; separation is much greater in the inflation forecast case (Figure 8, both panels; there is little observable distinction between the fixed and market interest rate cases), suggesting much higher forecast resolution. GDP growth forecasts (Figure 7) in fact show little separation of the CDF's after the shortest horizons, a result which mirrors the low 'content horizon' on U.S. and Canadian GDP growth point forecasts reported by, for example, Galbraith 2003 and Galbraith and Tkacz 2007: that is, forecasts of GDP growth generally do not improve markedly on the simple unconditional mean beyond about one or two quarters into the future, and the Bank of England forecasts appear to reflect the general difficulty of this problem. However, for GDP forecasts there is some observable distinction between the fixed and market interest rate cases; the fixed cases show somewhat higher resolution in at short horizons.

These differences are confirmed by the numerical results on scaled resolution presented in Table 2. Scaled resolution (recall that this estimate is bounded to the [0,1] interval) in GDP growth forecasts is low even at short horizons, and approximately zero at moderate and long horizons; by contrast, inflation forecasts show substantial resolution for several quarters. The market-interest-rate forecast for inflation shows the highest resolution.

6. Concluding remarks

By focusing our attention on particular probabilities derived from the fan charts, we have evaluated the performance of these density forecasts in problems of the type likely to be of interest to forecast users, as opposed to a general evaluation of whether the fan charts represent correct conditional densities; it is of course possible that a forecast density differs in some respects from the conditional density while nonetheless producing probability forecasts with some valuable features.

A number of interesting empirical results emerge. First, it is apparent even from the preliminary graphical results presented here that the predicted densities do not entirely conform with the true conditional densities. We then map these densities onto particular probabilistic forecasts for evaluation of the calibration and resolution. For inflation forecasts, deviations from correct calibration appear to be small, although nonetheless statistically significant at a number of forecast horizons. GDP growth forecasts produce much larger estimated deviations from correct calibration: that is,

predicted probabilities of GDP falling below our threshold are in many cases far from our estimates of the true conditional probabilities. However, only a subset of the observed deviations are statistically significant, perhaps because of the limited sample size.

Resolution falls rapidly with forecast horizon, and is higher for inflation forecasts.

These results, particularly at longer horizons, reflect differences in inflation and GDP growth observed in other contexts: the usefulness of conditioning information allowing us to make forecasts appears to decay much more quickly for GDP growth, and the persistence in the data is much lower. Whether sufficient information exists to produce forecast densities for GDP growth which differ from the unconditional densities, at the longer horizons used by the Bank of England, is an open question.

References

Brier, G.W. (1950) "Verification of forecasts expressed in terms of probabilities." *Monthly Weather Review* 78, 1-3.

Brittan, E., P. Fisher and J. Whitley (1998) "The *Inflation Report* Projections: Understanding the Fan Chart." *Bank of England Quarterly Bulletin,* 30-37.

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

Clements, M.P. (2004) "Evaluating the Bank of England density forecasts of inflation." *Economic Journal* 114, 844-866.

Corradi, V. and N. Swanson (2006) "Predictive density evaluation." in Elliott, G., C. Granger and A. Timmerman, eds., *Handbook of Economic Forecasting,* North-Holland, Amsterdam.

Croushore, Dean and Tom Stark (2003) "A Real-Time Dataset for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics* 85(3), 605-617.

Diebold, F.X. and G.D. Rudebusch (1989) "Scoring the leading indicators." *Journal of Business* 62, 369-391.

Dowd, K. (2008) "The GDP fan charts: an empirical evaluation." *National Institute Economic Review* 203, 59-67.

Elser, R., G. Kapetanios, T. Taylor and T. Yates (2005) "Assessing the MPC's fan charts." *Bank of England Quarterly Bulletin,* 326-348.

Galbraith, J.W. (2003) "Content horizons for univariate time series forecasts". *International Journal of Forecasting* 19, 43-55.

Galbraith, J.W. and S. van Norden (2008) "The calibration of probabilistic economic forecasts." Working paper, CIRANO.

Galbraith, J.W. and G. Tkacz (2007) "Forecast content and content horizons for some important macroeconomic time series." *Canadian Journal of Economics* 40, 935-953.

Gneiting, T., F. Balabdaoui and A.E. Raftery (2007) "Probabilistic forecasts, calibration and sharpness." *Journal of the Royal Statistical Society Ser. B* 69, 243-268.

Hamill, T.M., J.S. Whitaker and X. Wei (2003) "Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts." *Monthly Weather Review* 132, 1434–1447.

Lahiri, K. and J.G. Wang (2007) "Evaluating probability forecasts for GDP declines." Working paper, SUNY.

Murphy, A.H. (1973) "A new vector partition of the probability score." *Journal of Applied Meteorology* 12, 595-600.

Murphy, A.H. and R.L. Winkler (1987) "A general framework for forecast verification." *Monthly Weather Review* 115, 1330-1338.

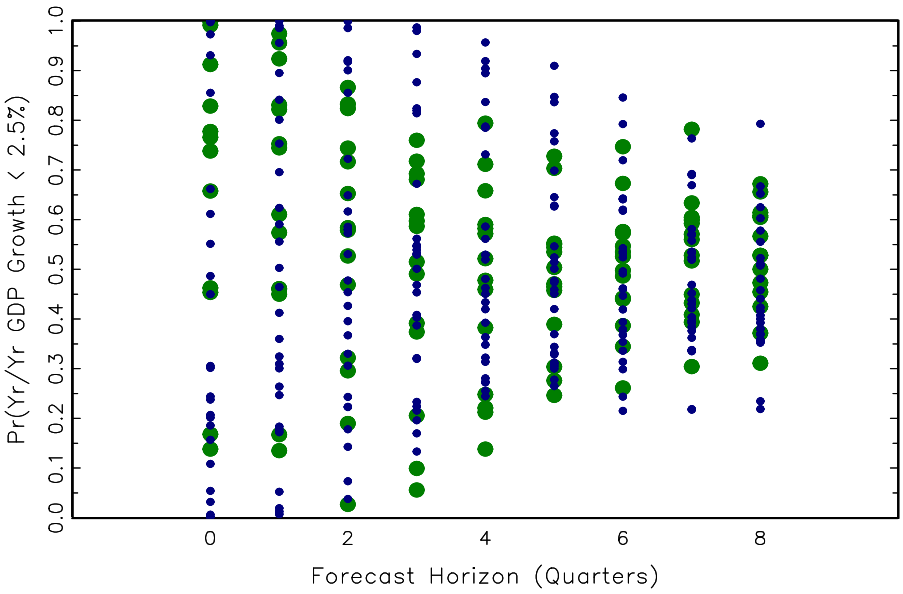
Orphanides, A. and S. van Norden (2005) "The reliability of inflation forecasts based on output gap estimates in real time." *Journal of Money, Credit and Banking* 37, 583-601.

Rudebusch, G. and J.C. Williams (2007) "Forecasting recessions: the puzzle of the enduring power of the yield curve." Working paper, FRB San Francisco.

Wallis, K.F. (2003) "An Assessment of Bank of England and National Institute Inflation Forecast Uncertainties." *National Institute Economic Review* 198, 64-71.

FIGURE 1
Implied probability forecasts from Bank of England fan charts

(i) GDP growth vs. threshold; fixed interest rates



(ii) GDP growth vs. threshold; market interest rates

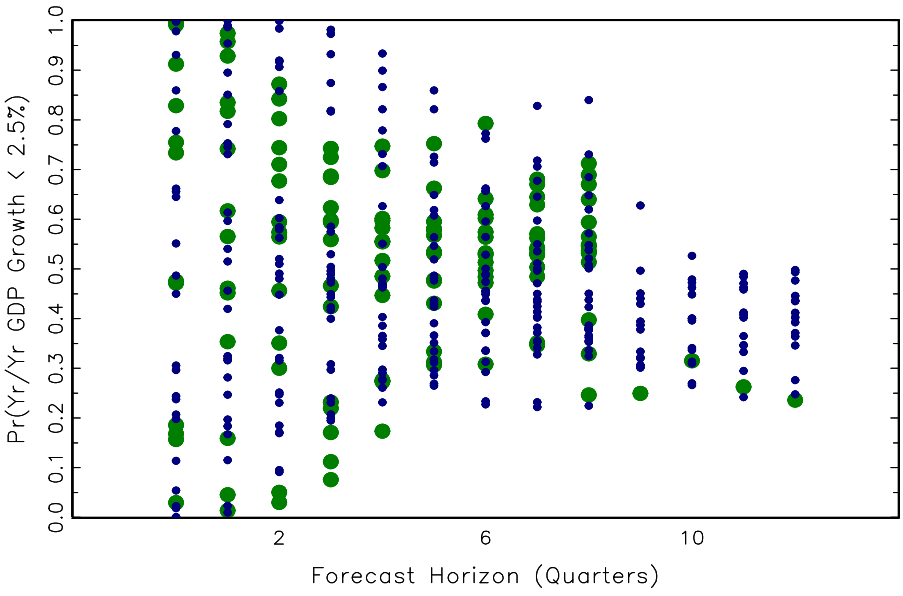
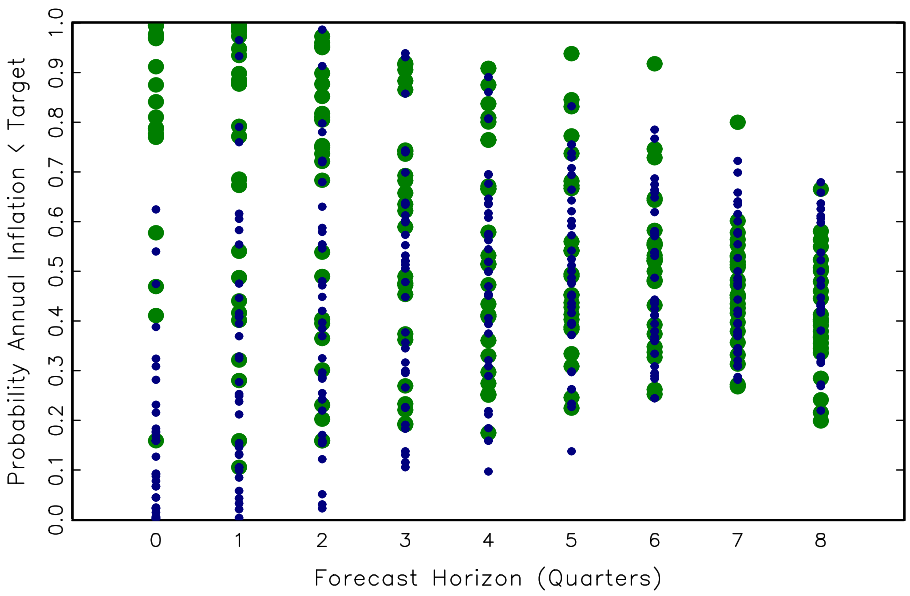


FIGURE 2
Implied probability forecasts from Bank of England fan charts

(i) Inflation vs. threshold; fixed interest rates



(ii) Inflation vs. threshold; market interest rates

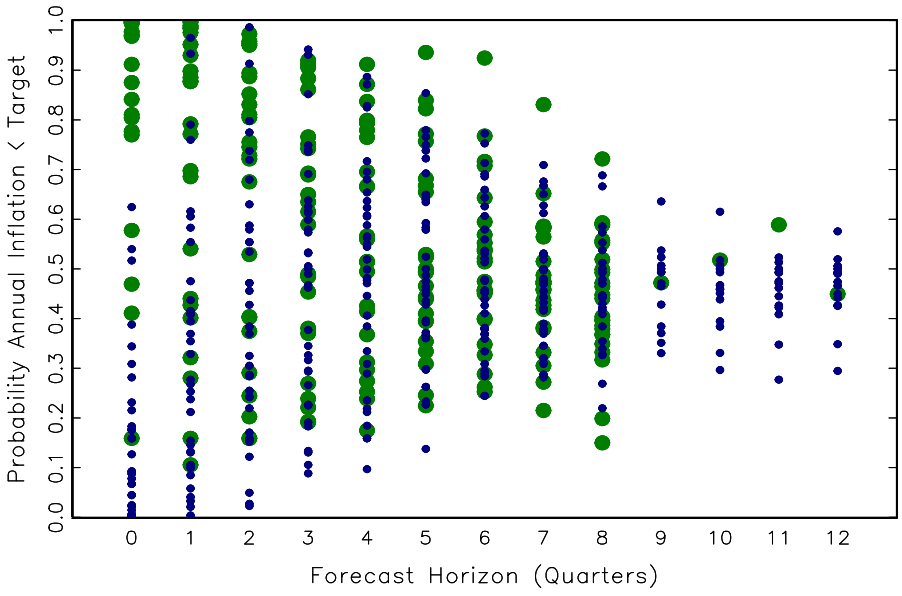
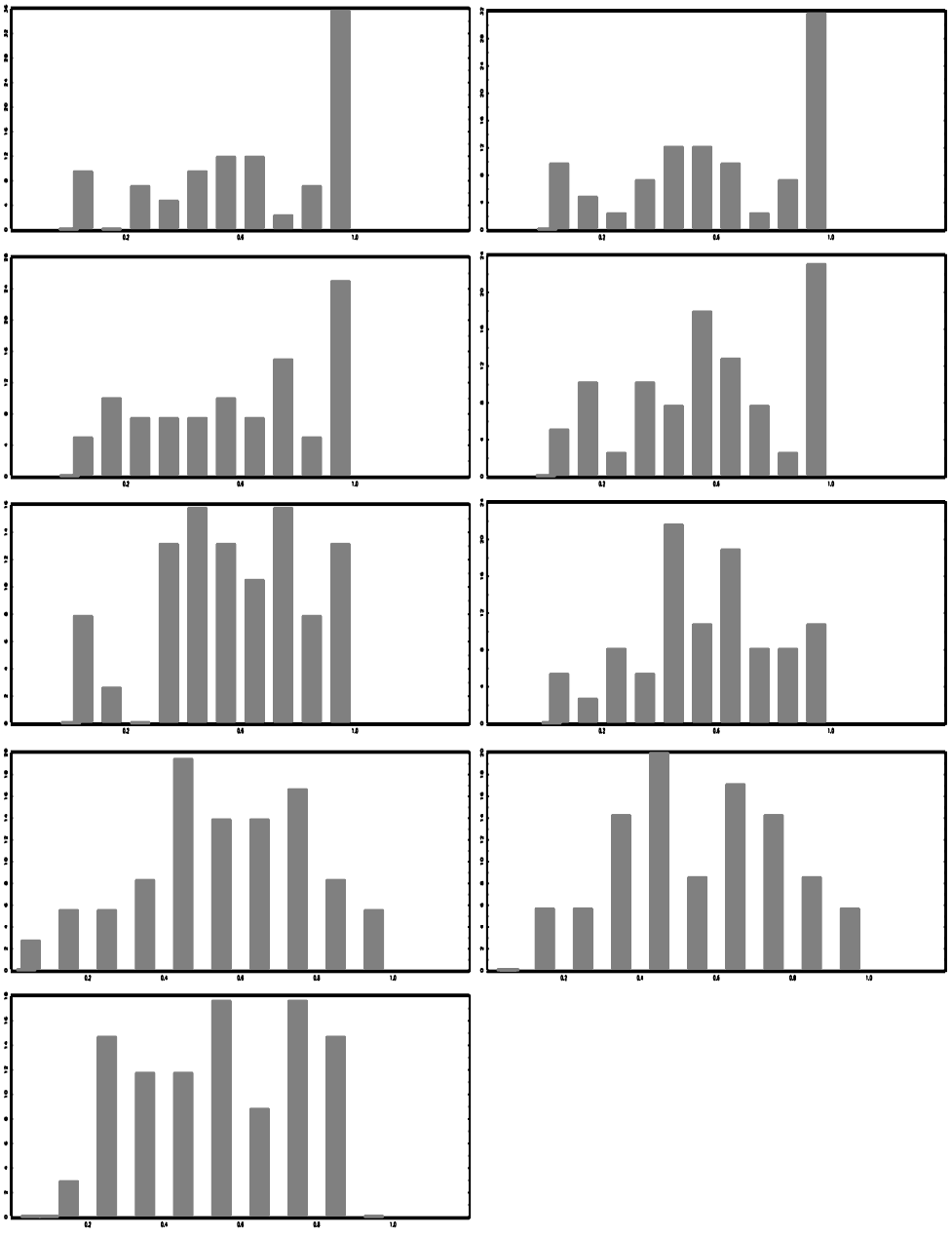


FIGURE 3(I)
Histograms of probability integral transforms[7]
GDP growth vs. threshold; fixed rates



[7]Horizons 0-8; bin width 0.1; proportionate frequency on vertical axis.

FIGURE 3(II)
Histograms of probability integral transforms
GDP growth vs. threshold; market rates

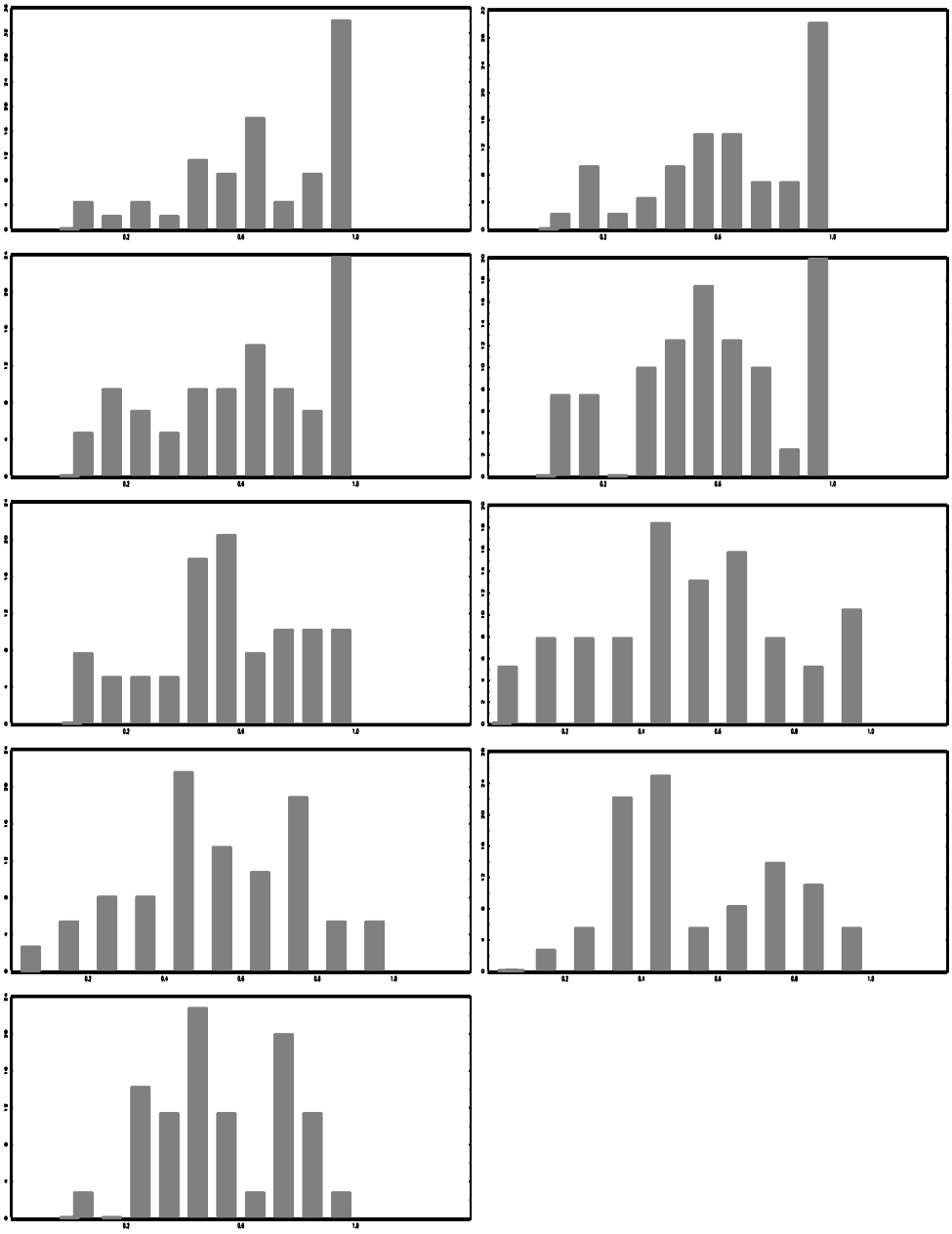
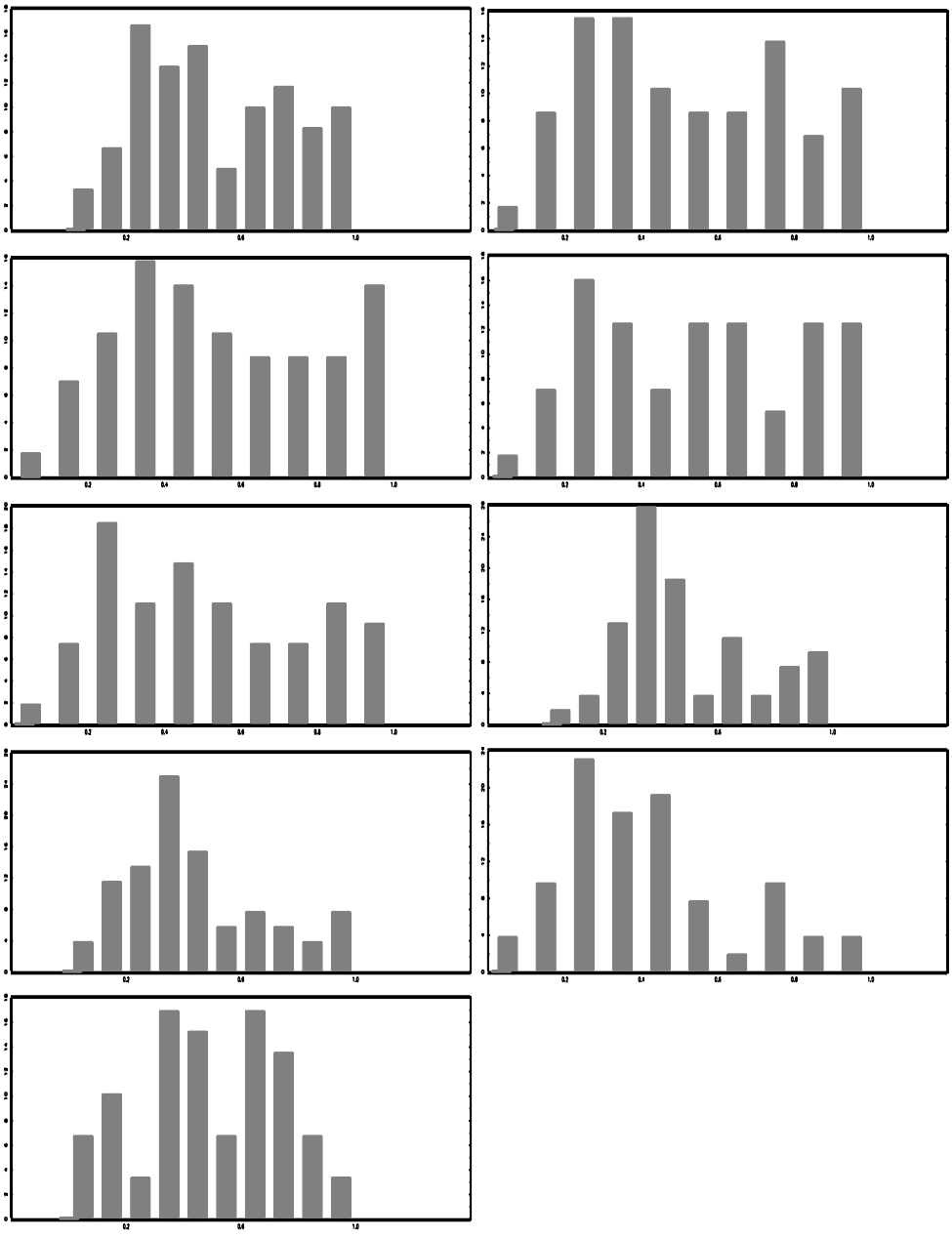


FIGURE 4(I)
Histograms of probability integral transforms[8]
Inflation vs. threshold; fixed rates



[8]Horizons 0-8; bin width 0.1; proportionate frequency on vertical axis.

FIGURE 4(II)
Histograms of probability integral transforms
Inflation vs. threshold; market rates

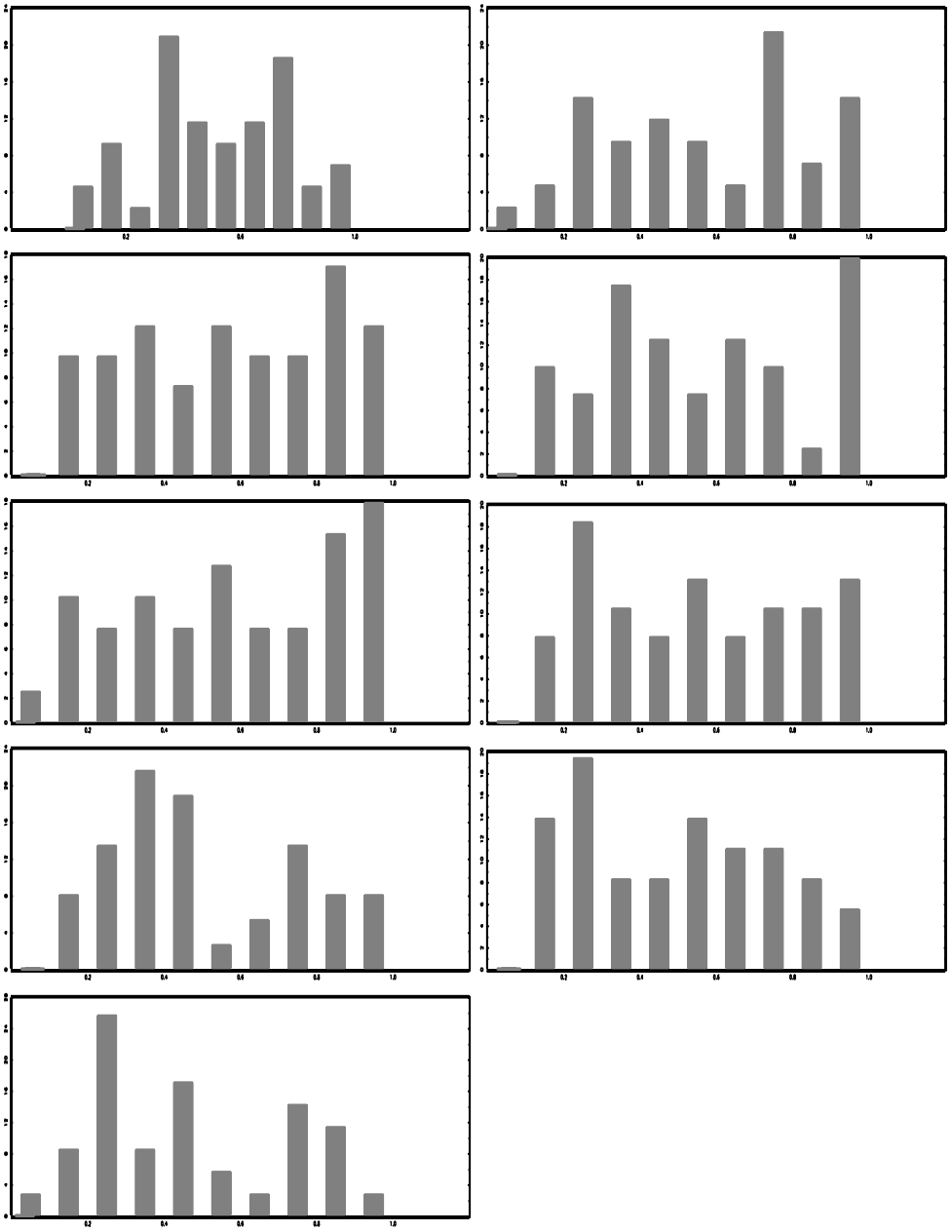
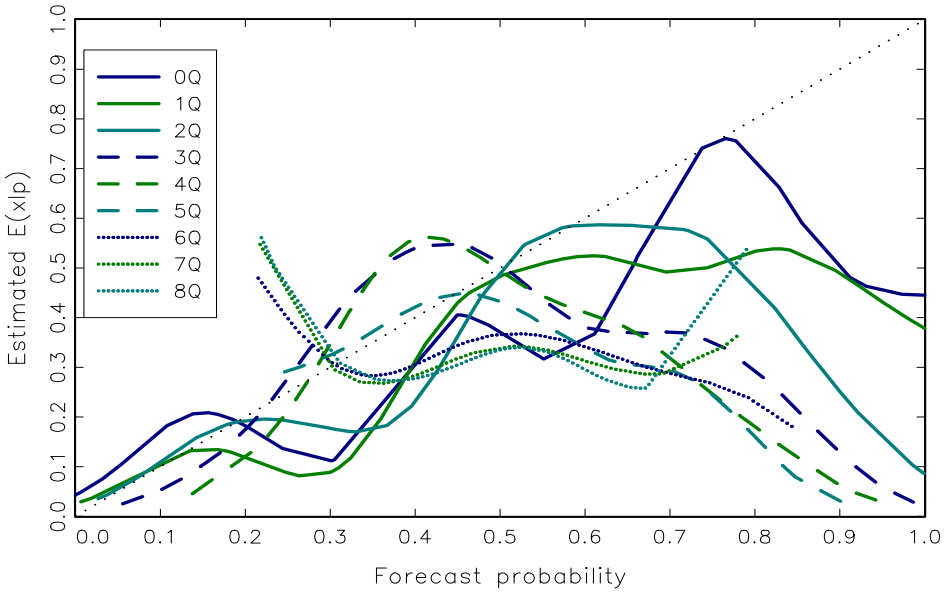


FIGURE 5
Calibration of GDP growth forecasts, 2.5% threshold

(i) 2008 Q2 vintage data, fixed interest rates



(ii) 2008 Q2 vintage data, market interest rates

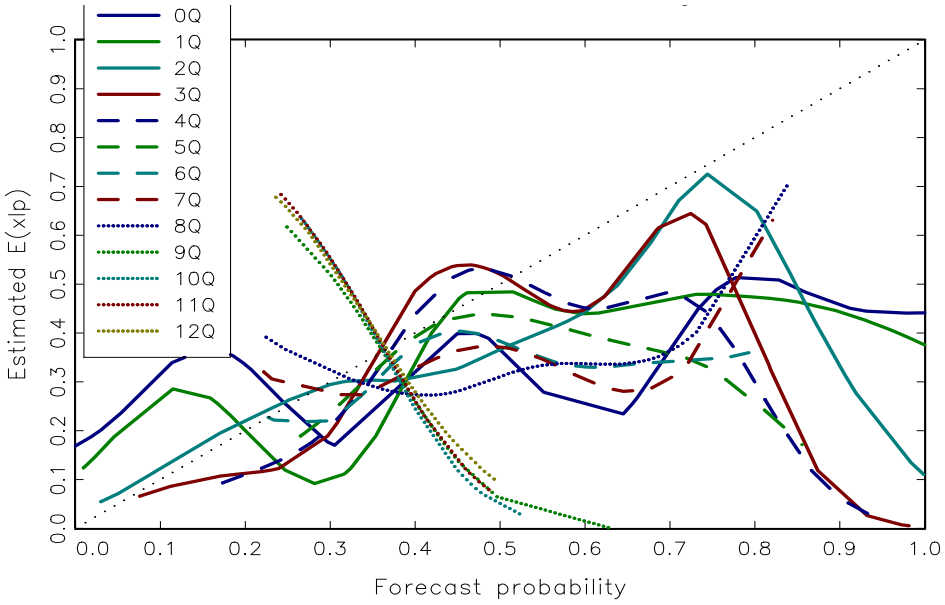
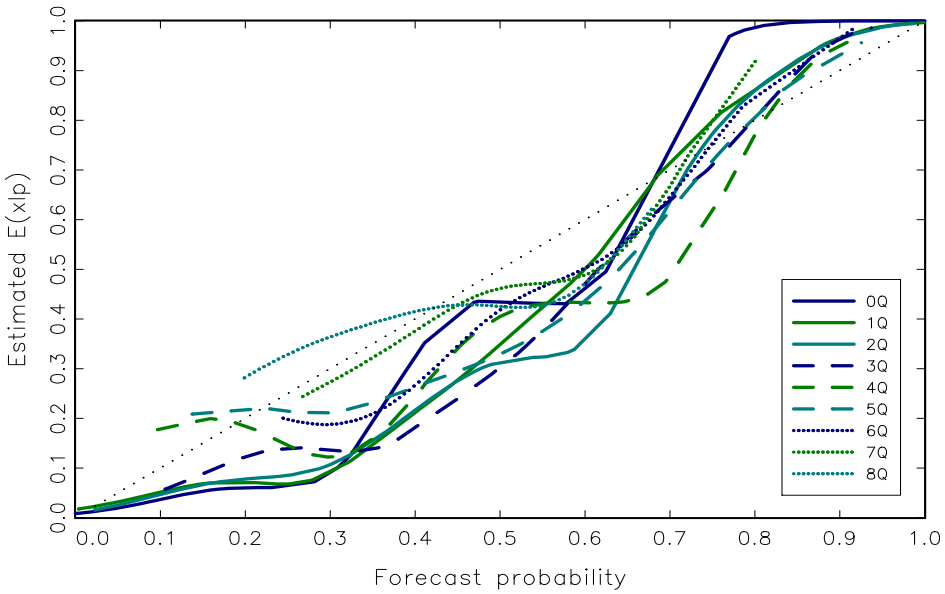


FIGURE 6
Calibration of inflation forecasts, target threshold

(i) 2008 Q2 vintage data, fixed interest rates



(ii) 2008 Q2 vintage data, market interest rates

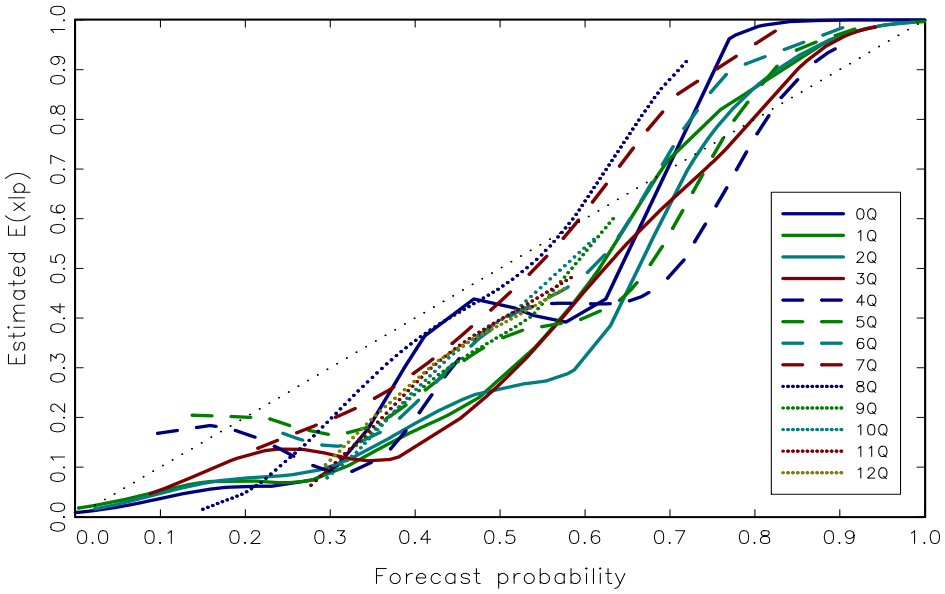
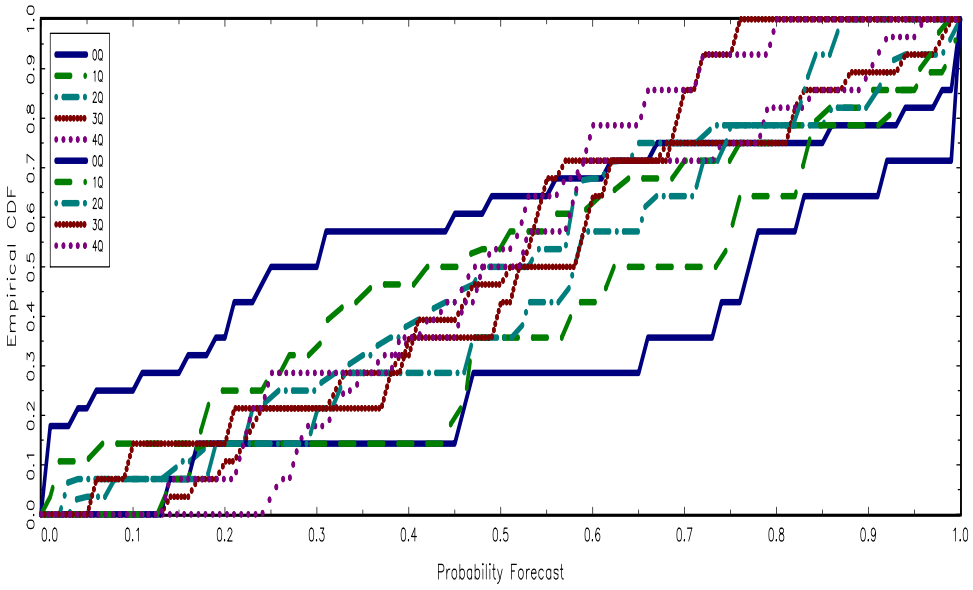


FIGURE 7
Empirical CDF's of implied probability forecasts from fan charts
Cases of GDP growth above/below threshold
(i) fixed interest rates



(ii) market interest rates

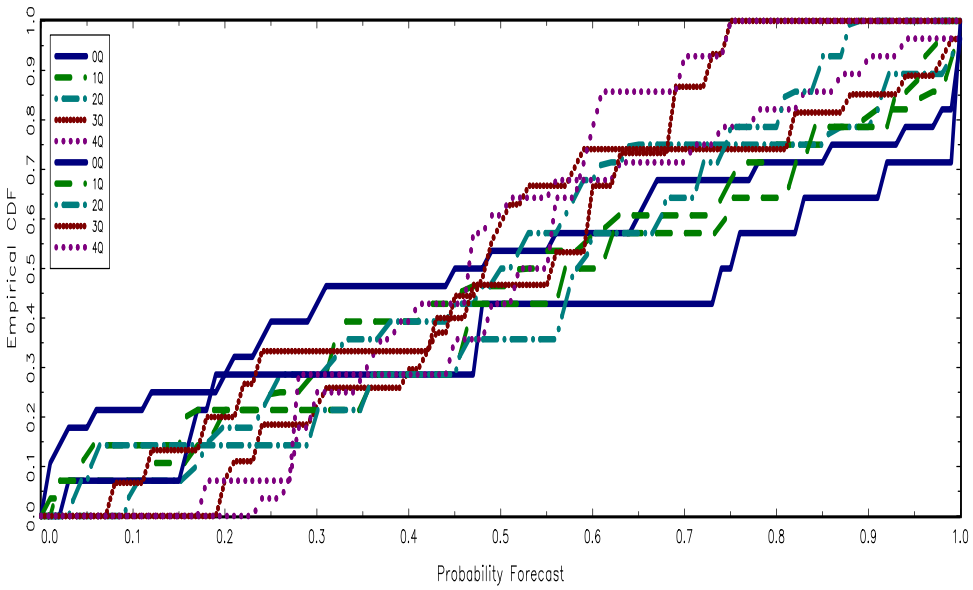
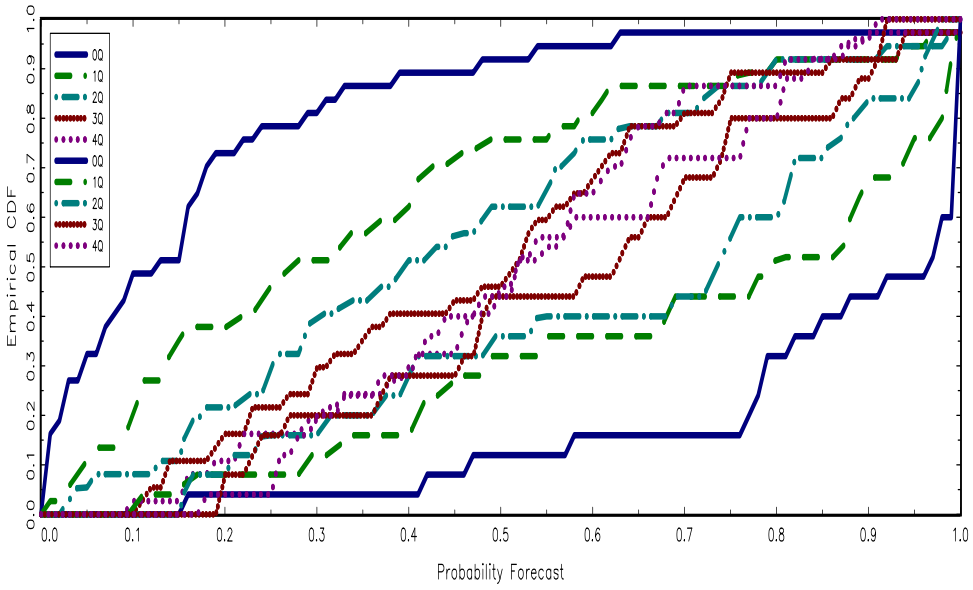


FIGURE 8
Empirical CDF's of implied probability forecasts from fan charts
Cases of inflation above/below threshold
(i) fixed interest rates



(ii) market interest rates

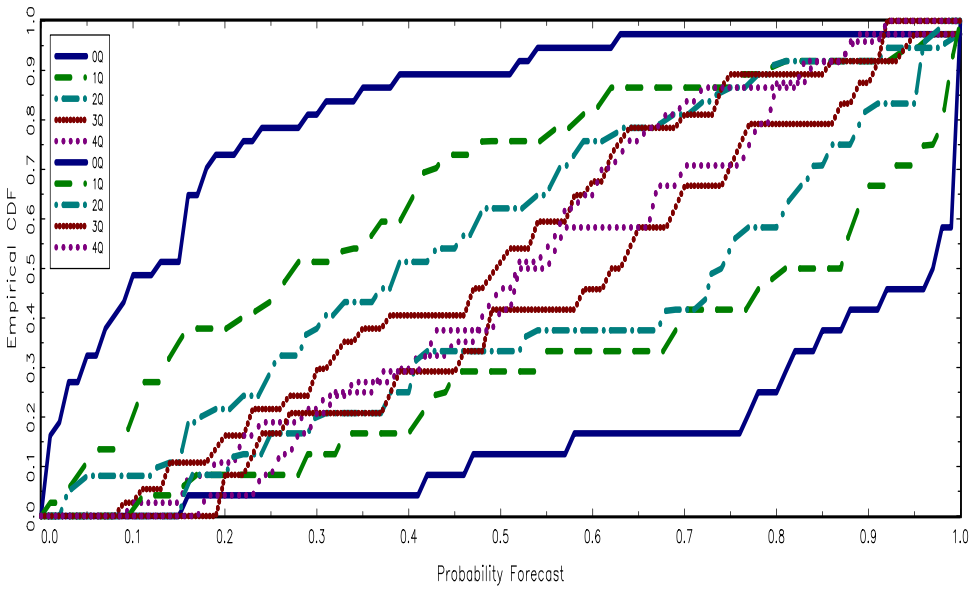


Table 1
p-values in linear test of calibration
$H_0 : a = 0.b = 1$ in $E(x|\hat{p}) = a + b\hat{p}$

Horizon	GDP, fixed	GDP, market	Infl., fixed	Infl., market
0	0.036	0.004	0.012	0.021
1	0.037	0.018	0.021	0.014
2	0.020	0.041	0.013	0.010
3	0.003	0.096	0.078	0.063
4	0.000	0.169	0.356	0.277
5	0.001	0.297	0.408	0.308
6	0.010	0.334	0.319	0.091
7	0.106	0.396	0.579	0.001
8	0.252	0.429	0.962	0.000

Table 2
scaled resolution measure ($\in [0, 1]$)

Horizon	GDP, fixed	GDP, market	Infl., fixed	Infl., market
0	0.191	0.054	0.798	0.791
1	0.155	0.087	0.600	0.628
2	0.162	0.108	0.498	0.526
3	0.124	0.193	0.350	0.371
4	0.129	0.118	0.217	0.245
5	0.045	0.032	0.169	0.209
6	0.011	0.013	0.134	0.207
7	0.015	0.017	0.055	0.142
8	0.022	0.022	0.017	0.125