

THE MEASUREMENT AND CHARACTERISTICS OF PROFESSIONAL FORECASTERS' UNCERTAINTY

GIANNA BOERO, JEREMY SMITH AND KENNETH F. WALLIS*

Department of Economics, University of Warwick, Coventry, UK

SUMMARY

Several statistical issues that arise in the construction and interpretation of measures of uncertainty from forecast surveys that include probability questions are considered, with application to the Bank of England Survey of External Forecasters. Substantial heterogeneity of individual forecasters' uncertainty is found, together with significant persistence in their relative uncertainty, which is a new finding in professional forecast surveys. It is an individual characteristic akin to the individual optimism and pessimism already established in the literature on point forecasts; the latter is also found in the current dataset, now in a bivariate sense with respect to joint inflation and output growth point forecasts. Whether disagreement among point forecasts is a useful indicator of uncertainty is shown to depend on the underlying macroeconomic environment. Copyright © 2014 John Wiley & Sons, Ltd.

Received 1 February 2013; Revised 16 November 2013

1. INTRODUCTION

Macroeconomic risk and uncertainty is an abiding preoccupation of policy makers, analysts and researchers, who draw on a wide range of indirect measures, indicators and proxy variables when assessing the current economic conjuncture and future economic prospects. In contrast to these proxy variables, a more direct assessment is provided by measures of uncertainty obtained from survey data, in particular from surveys of professional forecasters, which are receiving increasing attention. In a recent article Soderlind (2011), for example, shows how measures of inflation uncertainty based on survey data are a useful indicator of inflation risk premia, thereby helping to understand the evolution of 'break-even inflation' (the difference between nominal and real interest rates). Most of his empirical analysis is based on US data, which include the US Survey of Professional Forecasters managed by the Federal Reserve Bank of Philadelphia, although he also considers two further countries that have survey data on inflation uncertainty. One is the UK, for which he uses a series constructed from the Bank of England Survey of External Forecasters in one of our earlier articles (Boero *et al.*, 2008); the second is the euro area, for which he uses the European Central Bank's Survey of Professional Forecasters. Soderlind (2011) also provides useful references to the related literature, some of which reappear below. Almost all of the empirical literature is based on the US data, whereas the empirical work in the present article continues our analysis of the Bank of England dataset, which is updated annually.

The raw data on uncertainty in these quarterly surveys comprise the responses of individual members of a panel of forecasters to questions asking for their personal probabilities that the value of the variable of interest (inflation, output growth, etc.) in a specified future period will lie in each one of a number of pre-assigned intervals. In this article we consider several statistical issues that arise in the construction and interpretation of measures of uncertainty from density forecasts reported in this form, and discover some interesting properties of the resulting measures.

The Bank of England Survey of External Forecasters covers a sample of City firms, academic institutions and private consultancies, predominantly based in London. Their identities are not known

* Correspondence to: Kenneth F. Wallis, Department of Economics, University of Warwick, Coventry CV4 7AL, UK.
E-mail: K.F.Wallis@warwick.ac.uk

to us, although an identification number allows us to track individual responses over time. In the beginning, in 1996, the quarterly survey asked for point and density forecasts of inflation on a fixed-target basis, namely the end-quarters of the current year and the next year, as is the long-established practice of the US SPF. In contrast, a fixed-horizon question, asking for forecasts 2 years ahead, was added in 1998, when GDP growth forecast questions were also added. May 2006 saw the two fixed-target questions replaced by fixed-horizon questions, 1 year ahead and 3 years ahead, so since that time we have quarterly series of 1-, 2- and 3-year-ahead forecasts, for both variables, and these are the data studied in this article. The dataset extends to November 2012 and so comprises 27 surveys. The format of the questionnaire is illustrated by the recent inflation question shown in Figure 1. For a forecast target period of quarter q , year T , the target variable is the percentage increase in the Consumer Price Index since quarter q , year $T - 1$.

A summary of survey results is published in each quarterly Bank of England *Inflation Report*, most of whose content comprises the economic analysis and forecasts of the Monetary Policy Committee; we refer to the surveys by the month in which publication occurs (February, May, August, November), although the survey is undertaken in the last few days of the preceding month. Since the latest inflation and growth data available to the forecasters at the time relate to the previous quarter, we refer to the 1-, 2-, and 3-year-ahead forecasts as having horizon h equal to 5, 9 and 13 quarters respectively. The summary report includes the survey average density forecasts, i.e. the average of the respondents' probabilities in each interval. Each published table or chart notes the number of responses on which it is based; this is typically in the low twenties, and varies by forecast horizon. This survey, like most others, thus faces the problems of item non-response as well as complete non-response.

The period since the questionnaire redesign in May 2006 includes the recent financial crisis, with more action in the data than in the period covered by our 2008 article, when the range of the quarterly observations of the then-target annual RPIX inflation rate was 1.9–3.2%. The Governor of the Bank of England famously described the earlier experience as *non-inflationary consistently expansionary*, or 'nice' (King, 2003), whereas recent experience has been rather different, and has led to occasional changes in the questionnaire, as discussed below.

The remainder of the article is organized as follows. Section 2 considers the statistical framework for the measurement and analysis of uncertainty, and compares several approaches to the construction of measures of uncertainty. For our preferred measure, based on fitted normal distributions, we observe substantial heterogeneity of individual forecasters' uncertainty. In Section 3 we establish that there is significant persistence in individual relative uncertainty, using a statistical procedure based on rank orderings of individual uncertainty. Although relatively high or low uncertainty cannot be

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 2011 Q4, 2012 Q4 and 2013 Q4. 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			
	2011 Q4	2012 Q4	2013 Q4
<0%			
0.0% to 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		
2011 Q4	2012 Q4	2013 Q4

Figure 1. Bank of England questionnaire, November 2010 survey, inflation question

characterized as pessimism or optimism, these terms have been applied to similar findings of persistence in the relative level of individual point forecasts, using different statistical procedures, on different datasets. Accordingly we turn to the reported point forecasts in Section 4, and replicate this finding using our statistical procedure on the present dataset. Disagreement among individual point forecasts has been much discussed as a possible proxy for uncertainty in the absence of a direct measure of uncertainty, and we return to this question in Section 5, finding that the recent more turbulent period is rather more informative than the 'nice' period. It also illustrates the limitations of the published survey average density forecast as an indicator of uncertainty. Section 6 concludes.

2. MEASURING UNCERTAINTY

2.1. The Use of Probability Distribution Functions

Any graphical presentation or statistical analysis of the responses to a question such as that shown in Figure 1 rests, implicitly or explicitly, on an assumption about an underlying probability distribution for the variable in question. For example, the traditional method of calculating mean and variance applies standard formulae for discrete distributions using representative values of the variable in each interval. Such values are usually taken to be the midpoints of the intervals, adding an assumption about the highest and lowest intervals, which are open: these are commonly treated as closed intervals of the same width or twice the width of the interior intervals. Thus the traditional approach treats the distribution as if all probability mass is located at the interval midpoints.

The underlying variable is continuous, however, and the simple assumption that underlies the conventional graphical presentation in histogram form is that the probability distribution is uniform within each interval or histogram 'bin'. This assumption is also conventional in forecast evaluation exercises, where probability integral transforms and log scores of the outcomes are calculated by linear interpolation within the relevant interval. In effect, forecasters have reported a small number of points on their forecast cumulative distribution functions (CDFs), and the researcher is assuming that the CDFs are piecewise linear. In estimating moments, this assumption gives the same mean but increases the variance from the value given by the traditional approach by one-twelfth of the squared bin width.

Typical macroeconomic density forecasts are unimodal, however, suggesting that more of the probability in each bin is located closer to the centre of the distribution. The researcher may also prefer to match the continuity of the variable by assuming a continuous CDF. This has motivated estimation of the variance by fitting normal distributions, which do not require the range of the variable to be restricted by an assumed closure of the open-ended highest and lowest intervals. This is applied to the individual and average US SPF density forecasts by Giordani and Soderlind (2003), whose Figure 3 shows that, for the survey average histograms, the fitted normal distributions imply smaller variances than those given by the traditional method.

As an alternative to the normal distribution, Engelberg *et al.* (2009) fit generalized beta distributions to the individual US SPF density forecasts. The generalized beta distribution is an ordinary beta distribution that is scaled to have support (L, U) , where L and U are the bounds on the variable of interest. If the forecaster has placed non-zero probability in one or both of the open intervals, then closure assumptions are again required. With parameters $p > 1$, $q > 1$ the beta distribution matches the unimodal character of macroeconomic density forecasts, and in addition is very flexible (Johnson *et al.*, 1995, ch. 25). If the implied point forecast is the focus of attention, for example, then the beta distribution allows different mean, median and mode, each of which can be defended as an appropriate point forecast under different loss functions. As is the case with the normal distribution, non-zero probabilities are needed in at least three bins to allow fitting to proceed; on occasions when a forecaster uses only two intervals, which are invariably adjacent, Engelberg *et al.* (2009) fit symmetric triangular distributions. The same general approach, based on the beta distribution, is used by Bruine de Bruin

et al. (2011) in a study of consumers' uncertainty about future inflation. We compare the normal and beta approaches to the measurement of uncertainty in Section 2.3.

2.2. Uncertain Uncertainty

Before turning to the empirical comparison of these two distributions, we note two reasons for exercising caution in interpreting any probability distribution as a perfect representation of forecast uncertainty. One is that the relatively few probability assessments provided by the forecaster are not sufficient to specify a unique probability distribution, while any distribution chosen by the researcher expresses uncertainty about the variable in more detail than the forecaster has provided. It is possible, however, that the quantity of interest derived from the distribution, which in our present case is a measure of its dispersion, is robust to changes in the chosen distribution, and this is our focus of attention below.

The second reason for caution is that it is difficult for forecasters to give precise numerical values for their subjective probabilities. This is immediately clear on looking at forecast survey data, where forecasters' uncertainty about their subjective probabilities is demonstrated by their widespread yet varying use of round numbers.

In statistical reporting, the level of rounding implicitly conveys information about the quality or accuracy of the data, and so aids interpretation of the data. Similarly, individual forecasters' level of rounding of their reported probabilities conveys information about the subjective uncertainty inherent in their probability assessments, which varies across individual forecasters. The overall extent of rounding is indicated in Table I, which shows the proportion of non-zero percentage probabilities of different numerical characteristics observed in the complete dataset: all respondents, all histogram bins, both variables, and three forecast horizons, almost 18,000 numbers in total. It is seen that almost two-thirds of all bin probabilities are even or odd multiples of 5. Multiples of 10 are used in every bin of the histogram in approximately 10% of the individual forecasts. More commonly, round numbers are used in the centre of the distribution and smaller numbers in the tails, which also include most of the non-integer values recorded in Table I. There is thus considerable variation in the treatment of uncertain probability assessments, both in the pattern of rounding across the distribution and in the extent of rounding by individual respondents. Nevertheless, the extent and pattern of rounding by individual survey respondents have strong similarities across the three forecast horizons, the two variables and over time. Disaggregating Table I by forecast horizon, for example, shows essentially the same distribution at horizons 5, 9 and 13: it is well known that uncertainty about point forecasts increases with forecast horizon, but it appears that uncertainty about forecast uncertainty does not, at least insofar as it is reflected in this aspect of forecast practice and/or forecaster behaviour, which is worthy of further investigation.

Engelberg *et al.* (2009, Appendix) assess the sensitivity to the rounding of probabilities of the conclusions of their non-parametric analysis of the bounds on the subjective means, medians and modes implied by the density forecasts in the US SPF. Manski and Molinari (2010) study the rounding of responses to questions that ask members of the public to state their percent chance that some future event will occur, concerning their health or retirement, for example. The availability of responses to

Table I. The numerical character of reported probabilities

Reported percent probability	Percentage of cases
Multiple of 10	34.2
Otherwise multiple of 5	30.6
Other integer	31.5
Non-integer	3.8

several such questions allows an individual's rounding practice to be inferred, so that an interval can be attached to the response and the resulting data appropriately analysed. Extension of these ideas to our present context remains on the research agenda.

In contrast to percentage probabilities stated as round numbers, the reported probabilities of a small number of survey respondents appear to be calculated from a known probability distribution. The normal distribution, with possible extensions to asymmetry, is a convenient and popular choice among forecasters who publish density forecasts. The distribution is typically centred on the associated point forecast, with variance calibrated to recent past point forecast errors, possibly with judgmental adjustment. In such cases the estimation of forecast uncertainty, to which we turn next, simply amounts to the recovery of the parameter values used by the forecaster.

2.3. Choosing a Measure

We first report the results of a comparative study of variance estimates based on the normal and beta distributions. In each case the parameters are estimated by fitting the cumulative distribution function to the forecast CDF by least squares, and we begin by discussing the results for the inflation forecasts.

To fit either distribution it is required that the forecaster has assigned non-zero probabilities to at least three bins. Of the available 1749 individual forecasts of inflation this requirement excludes 36 cases, in which only two bins are used, and we adopt the triangular distribution of Engelberg *et al.* (2009) in these cases. In the great majority of the remaining cases, including most of the three-bin cases, the histograms plotted assuming a constant bin width are unimodal on an interior bin, and the two distributions have very similar goodness-of-fit and give very similar estimates of the variance, which are smaller than the estimate given by the traditional calculation of moments. The exceptions to this generalization allow some discrimination between the distributions, however, which is partly related to shortcomings in the questionnaire design. To set the scene for the following discussion, we show in Figure 2 the real-time data on each variable, together with the range covered by the interior closed intervals of the questionnaire at the time.

The seven-bin inflation question shown in Figure 1 has been in use since the February 2009 survey. Previously, six bins were specified, with four half-percentage-point bins covering the range 1–3%, and open bins above and below this range. During 2008 there was much public discussion of the possibility of below-target inflation, if not deflation, and in the survey report published in the November 2008 *Inflation Report* the average percentage probability in the lower open bin was, for the first time, in double digits, this occurring at all three forecast horizons. We believe that this prompted the division of the lower bin by the survey managers, to give the new format shown in Figure 1. However inflation prospects changed in 2010, and we began to see double-digit survey average percentage probabilities in the upper open bin, approaching 20% in the 1-year-ahead forecasts in August and November 2010. But no change to the questionnaire resulted. The consequence of these two episodes is that there are several individual forecast histograms which, when we construct them by replacing the open bins with closed bins of the same width as the adjacent interior bin, have a U-shaped or J-shaped appearance. The beta distribution can match these shapes, with parameters less than 1, but we believe that restricting the range of the variable in this way distorts the forecasters' intentions, since all the individuals involved in these cases have submitted forecasts with a single interior mode on other occasions. Changing the support of the beta distribution can moderate some of these cases, but such changes to the range of the variable are arbitrary. On the other hand, the normal distribution imposes its unimodal shape in a manner determined by the available interior observations, and this is our preferred solution. To the objection that this has an infinite range, we recall the traditional rule-of-thumb that the range of the normal distribution can be taken to be six times its standard deviation: even a forecaster rounding to the nearest 0.5% would report a probability of zero outside this range.

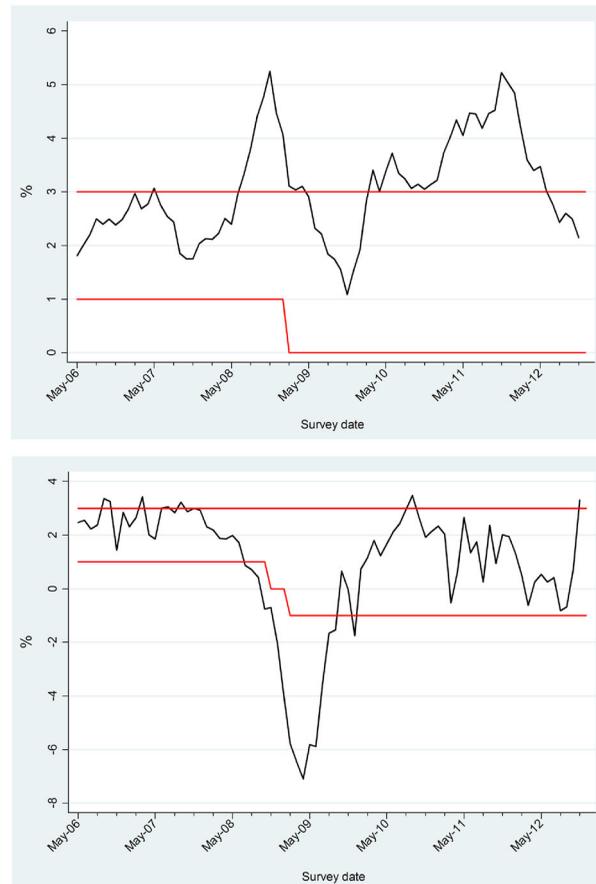


Figure 2. Real-time data and the range of the closed intervals. Upper panel: CPI inflation; lower panel: GDP growth

As an illustration, we choose our most extreme three-bin case, which is an individual 1-year-ahead forecast in the August 2010 survey, with percentage probabilities of inflation in the ranges 2–2.5%, 2.5–3% and >3% of 5, 5 and 90 respectively. How to choose a sensible upper bound that would allow a beta distribution with an interior mode to be fitted is an open problem. However, the normal distribution whose 5th and 10th percentiles are respectively 2.5 and 3.0 is an acceptable representation, with mean equal to 4.77 and standard deviation equal to 1.38, which is our preferred measure of uncertainty. We note that this respondent's reported central projection of CPI inflation was 4.0%, the highest among all survey respondents, and that the inflation outcome in 2011:Q3 was 4.7%. This example is also extreme in another sense, having the largest difference between the fitted mean and the reported point forecast. In all other similar cases, with a substantial probability assigned to an open interval, the point forecasts and the means of the fitted normal density forecasts are closer together, with the absolute difference ranging from zero to 0.55.

Turning to the GDP growth forecasts, we find further limitations in the design of the question. Over the period to August 2008 four bins were specified (<1%, 1–2%, 2–3%, >3%), but as recession fears increased the survey average probability observed in the lower bin increased, reaching 38% for the 1-year-ahead forecasts in August 2008, so this bin was divided for the November 2008 survey, and again in February 2009, since when six bins have been specified (<–1%, –1–0%, 0–1%, 1–2%,

2–3%, >3%). At the upper end of the range, with a long-run trend growth rate above 2% per annum, it is no surprise that the survey average probability of growth exceeding 3% in the 3-year-ahead forecasts is often in excess of 25%, but no comparable changes to the configuration of the bins have been made. The consequence is that we have more individual forecasts with J-shaped histograms in the growth forecasts, appearing when one or other open bin is closed, than in the inflation forecasts. Moreover, the use of wider interior intervals results in more cases in which there are only two bins to which non-zero probabilities have been assigned by the forecaster (67 out of 1741 available forecasts); also seven cases with 100% probability assigned to a single interior bin. For the two-bin cases we continue with the variance estimate from the triangular distribution of Engelberg *et al.* (2009), as above, and for the one-bin cases we set the variance equal to $\frac{1}{24}$, assuming a symmetric triangular distribution with unit support. The general results of a comparison between the beta and normal distributions for the GDP growth forecasts are otherwise very much in line with those obtained for the inflation forecasts, described above, but there are more exceptional cases for which the beta distribution is inappropriate. Hence our preferred uncertainty measure for both variables is the standard deviation of the fitted normal distribution, except in one- and two-bin cases, as noted above.

The resulting estimates of individual standard deviations are shown in Figure 3, whose six panels refer to the two variables and three forecast horizons. As a point of reference the solid line in each panel shows the median individual standard deviation, around which we observe substantial dispersion, implying substantial heterogeneity of individual forecasters' uncertainty. Note that the scale of the inflation panels is different from that of the GDP growth panels: the latter variable is generally considered to be more difficult to forecast, not least due to the problems caused by data revisions, whereas the CPI is never revised after first publication. The general level of uncertainty is lower at shorter forecast horizons, as expected. The median measure has a local peak in February 2009, which in some cases appears to signal a shift in level. The general spread of uncertainty measures also appears to have increased from that time. Figure 2 shows that February 2009 did not mark a turning point in either of the underlying series. Rather, these forecast measures reflect increased uncertainty about economic prospects as the global financial crisis spread, and central banks made unprecedented policy moves, with the MPC cutting UK bank rate by 3 percentage points between the November 2008 and February 2009 forecast surveys. Similarly substantial heterogeneity of forecast uncertainty, varying over time, is demonstrated over 35 years of the US SPF by Lahiri and Liu (2006, Figure 1).

3. THE PERSISTENCE OF INDIVIDUAL RELATIVE UNCERTAINTY

On moving to more systematic study of individual characteristics we immediately face the missing data problem. The Bank of England survey, like other forecast surveys and individual panel studies more generally, has experienced exit and entry of participants and sporadic non-response, both to the complete questionnaire and to items within it: the longer-horizon forecasts are more often missing than the 1-year-ahead forecasts. To avoid the complications caused by long gaps in the data we follow common practice in survey research and conduct our analysis of individual forecasters on a subsample of 'regular' respondents. In this article the subsample comprises the 17 respondents whose item response rate over the 27 surveys exceeds two-thirds. The overall subsample item response rate is 86.5%; over six questions and 27 surveys the number of available responses ranges between 13 and 17.

To study possible persistence in individual forecasters' relative uncertainty, we identify the regular respondents in Figure 3, and for each column in each panel of the figure rank them from the highest to the lowest uncertainty. In each panel we next calculate each forecaster's average rank over the (27 or fewer) surveys in which they appear. To illustrate possible persistence in relative positions within each panel of Figure 3, we select the five highest-ranked and the five lowest-ranked individuals, and show

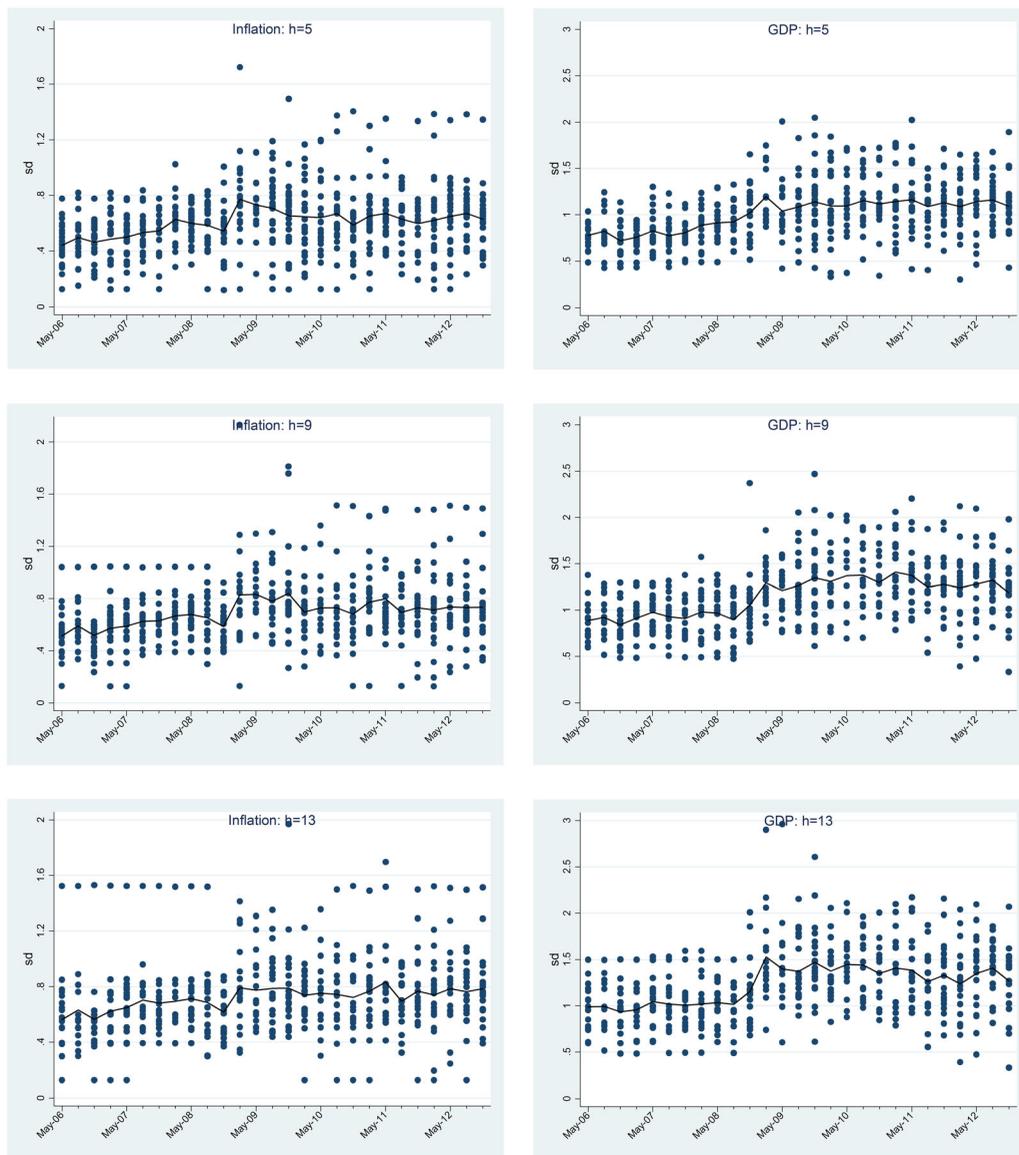


Figure 3. Spread of individual uncertainty measures, and median individual standard deviation. Left column: CPI inflation; right column: GDP growth; $h = 5, 9, 13$

their uncertainty measures in Figure 4, respectively with circles and plus signs. Note that missing observations imply that there are typically fewer than 10 points in each column; also these are not the same 10 individuals across the six panels, a question we return to below. To centre the spread, we retain the plots of the median individual uncertainty from Figure 3.

There is a very clear indication of persistence in relative forecast uncertainty in Figure 4, with circles tending to stay high and pluses tending to stay low in all six panels. Some of the obvious outliers in Figure 3 do not appear in Figure 4: most of these deletions are the uncertainty measures of one non-regular respondent; the others are due to a regular respondent whose average rank is not in the highest or lowest five. An eye-catching case is the single respondent whose uncertainty is the highest shown, at



Figure 4. Uncertainty measures of the five highest-ranked (circles) and lowest-ranked (pluses) regular respondents in each panel. Left column: CPI inflation; right column: GDP growth; $h = 5, 9, 13$

a constant level, over the first 10 surveys in our sample, in the inflation, $h = 9$ and 13, and GDP growth, $h = 13$ panels of Figure 4. In the next survey, November 2008, there was no response, then responses resumed, but no longer occupying the top position, and with a different pattern of rounding, suggesting that there had been a change of forecast personnel, models or methods (or all three) in this institution.

A statistical measure of the similarity over time of the rankings in each panel of these figures, and hence of the persistence of individual relative uncertainty, is provided by the Kendall coefficient of concordance (Kendall and Gibbons, 1990, ch. 6). Usually denoted W , this is defined as the ratio of the sum of squared mean deviations of the observed average ranks to its maximum possible value; thus $0 \leq W \leq 1$. With 17 regular respondents and no missing data, perfect agreement

of the rankings across all 27 surveys would give average ranks 1, 2, ..., 17 in some order, with sum of squared mean deviations equal to 408, which is the maximum possible value in this case. At the other extreme the average ranks all tend to equal 9 when rankings of individuals are purely random over time.

Whenever observations are missing, however, the maximum possible rank is less than 17. For each survey we rank the forecasts ignoring non-respondents, and individuals' average ranks are calculated over the occasions on which they responded. We calculate a revised maximum possible sum of squared mean deviations of average ranks conditional on the observed pattern of missing data in each case, and with the observed average ranks we obtain the results shown in Table II. To aid interpretation of these coefficients, we note that, under a null hypothesis of random rankings over time, with r rankings of n individuals and no missing data, $r(n-1)W$ is approximately distributed as chi-squared with $n-1$ degrees of freedom; hence the 95th and 99th percentiles of W in these circumstances, with $r=27$ and $n=17$, are 0.06 and 0.07 respectively. The coefficients in Table II thus indicate considerable stability over time in the relative level of individual forecasters' uncertainty, to a similar extent for both variables and all three forecast horizons. In order to pool these cases, we calculate the concordance between the six rankings given by the time-averaged scores in each case. The Kendall coefficient is 0.88; under the above null its 99th percentile for six rankings of 17 individuals is 0.33. The rankings of individual forecasters by their uncertainty levels are almost identical across the two variables and three forecast horizons we consider.

This strong evidence of persistence in individual forecasters' relative levels of uncertainty, as expressed in their subjective probabilities, is a new finding in the context of surveys of professional forecasters. On the other hand, in the context of consumer surveys, Bruine de Bruin *et al.* (2011) report a similar finding. A similar question about inflation expectations over the next 12 months, with six intervals covering the range -4% to 12% and open intervals above and below, is administered to members of RAND's American Life Panel, a group of individuals who have agreed to participate in occasional online surveys. Individual forecast uncertainty is measured by the inter-quartile range of a fitted beta distribution, following Engelberg *et al.* (2009), and this is seen to be highly persistent, with such persistence being mostly captured by permanent time-invariant idiosyncratic differences across individuals.

Returning to the US Survey of Professional Forecasters, we note that Lahiri and Liu (2006, p. 1206) report that 'forecast uncertainty shows some, but not a lot, of persistence' although this is a different notion of persistence from ours. This survey has a fixed-target 'end-year' design; thus in each individual/year cell of a panel data setup the regression of log forecast variance (quarterly) on its lagged value (and other variables) would be expected to have a coefficient that is less than one, since forecast uncertainty falls as the target period is approached. Autoregressive coefficients between 0.38 and 0.48 for the preferred estimators are the basis for the conclusion quoted above, but these are measuring the rate of decline of uncertainty as the forecast lead time falls, and not the persistence of uncertainty for a given lead time over successive surveys. Variance estimates are based on fitted normal distributions using the software of Giordani and Soderlind (2003), which assumes that a normal distribution can be fitted to a two-bin histogram, but this is incorrect, as discussed in the Appendix.

Table II. Measures of agreement over time between forecasters' rankings with respect to their uncertainty measures: Kendall coefficients of concordance

	$h=5$	$h=9$	$h=13$
CPI inflation	0.40	0.45	0.43
GDP growth	0.47	0.40	0.44

4. PERSISTENCE IN THE RELATIVE LEVEL OF POINT FORECASTS

In this section we use the statistical procedures developed above to replicate some existing findings of persistence in the relative levels of individual point forecasts, on which previous research has principally focused. For example, we note two studies based on data from the *Consensus Economics* service, which is a monthly survey of private sector forecasting bodies in a number of countries. Forecasts are collected for the current year and the following year, so forecasters eventually supply a series of 24 forecasts for each target year. Several researchers have made use of the individual point forecasts of GDP growth and CPI inflation. Batchelor (2007) finds persistent individual biases towards optimism or pessimism in GDP growth forecasts for the G7 countries at all horizons for the target years 1991–2004; for the inflation forecasts there is less consistency across countries and forecast horizons. Patton and Timmermann (2010) study the persistent behaviour of US forecasters by classifying their point forecasts in each year into three groups—high, medium, low—and studying the associated Markov transition matrix. This is done separately for ‘short-horizon’ (current year) and ‘long-horizon’ (following year) forecasts of GDP growth and CPI inflation: there is a significant tendency for forecasters to stay in the same group from one year to the next.

The individual point forecasts from the Bank of England’s survey are shown in the six panels of Figure 5, together with the survey mean point forecasts. The overall dispersion of individual forecasts around the mean is relatively small until early 2009, when there is considerable disagreement about the consequences of the crisis and the policy actions taken in response to it. To study possible persistence in the relative positions of individual forecasters’ point forecasts, we repeat the transition between Figures 3 and 4 seen above, again working with the 17 regular respondents. Thus we first delete non-regular respondents from Figure 5, and for each time period in each panel rank the regular respondents from the highest to the lowest point forecast. Next, in each panel we calculate each forecaster’s average rank over the periods for which they supplied a forecast. Finally we delete the forecasts of all respondents except those with the five highest and five lowest average ranks, and plot these with circles and pluses respectively, to obtain Figure 6.

As in Figure 4 there are clear indications of persistence, which in Figure 6 relate to the relative levels of point forecasts. Again there is an overall tendency for circles to stay high and pluses to stay low across each panel of Figure 6, which is more pronounced in the GDP growth forecasts, and less so in the inflation forecasts, in particular when the dispersion is relatively small. This visual assessment is confirmed by the Kendall coefficients of concordance presented in Table III. The persistence in the relative level of point forecasts of GDP growth is similar to that observed in the relative level of forecast uncertainty; for inflation, the persistence in relative point forecasts is less than that in the uncertainty measures. In comparing Figure 6 and Table III, we note that Table III is based on the rankings of point forecasts and not on the particular values of the forecast inflation or growth rates, hence in the inflation $h=13$ panel of Figure 6, for example, the highest and lowest observations in May–November 2009 receive the same weight in the concordance calculations as the highest and lowest observations in August 2007–May 2008, although the separation of these earlier observations is much smaller. The frequent interchange of position within this reduced spread contributes to the relatively small value of the coefficient in this case.

Again we pool the six cases represented in the panels of Figures 5 and 6 and the cells of Table III, and calculate the concordance between the six rankings implied by the time-averaged scores for each variable/forecast horizon combination, *inverting* the GDP rankings: the Kendall coefficient is 0.49. This is smaller than the value obtained in the previous section, comparing the rankings of forecasters by their uncertainty levels across the two variables and three forecast horizons; nevertheless it is still

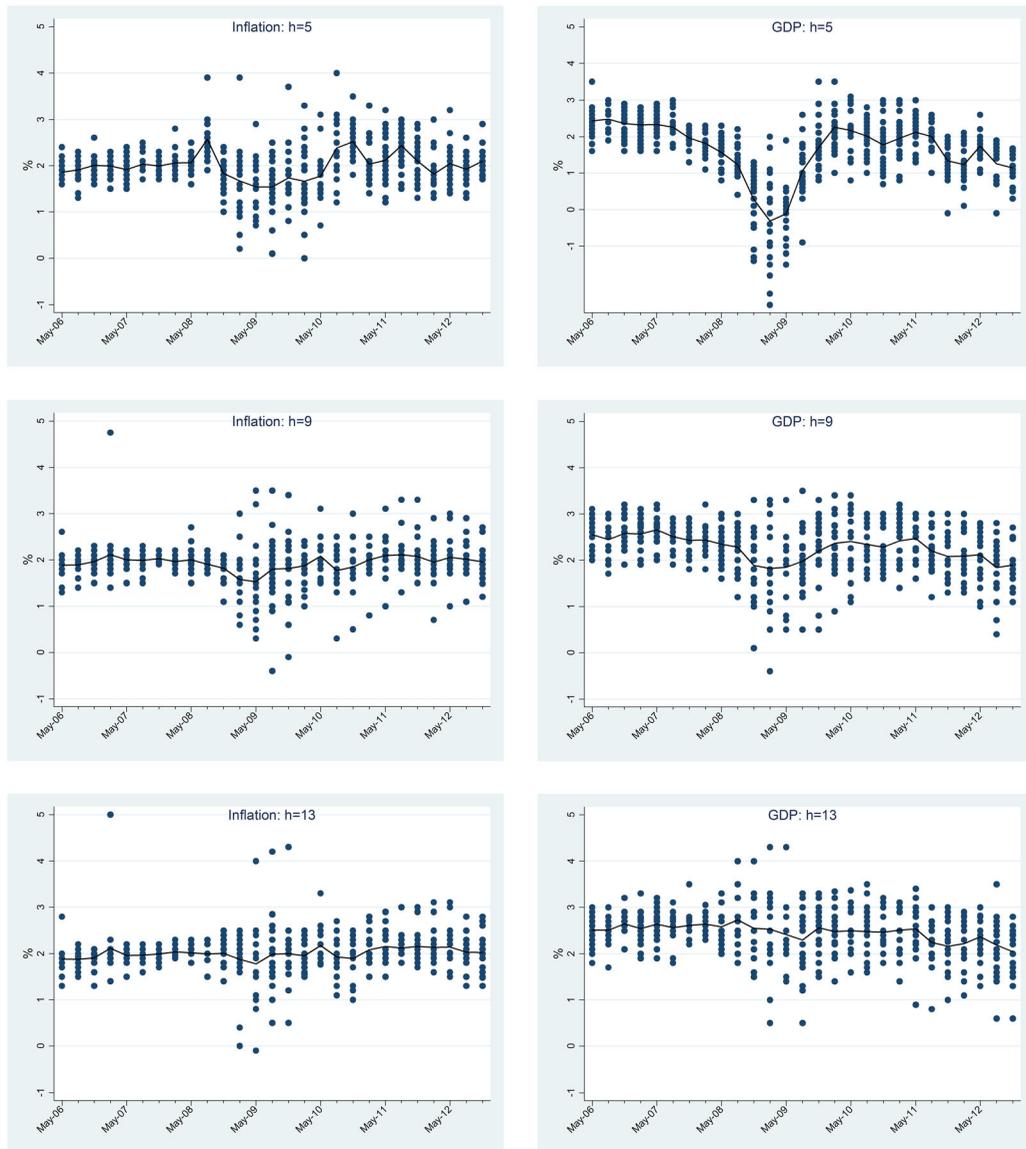


Figure 5. Spread of individual point forecasts and survey mean point forecasts. Left column: CPI inflation; right column: GDP growth; $h = 5, 9, 13$

well clear of the 99th percentile under the null, of 0.33, indicating strong similarity of these point forecast average rankings. Neither of the articles cited above considers a bivariate notion of optimism (forecasts of low inflation *and* high growth) and pessimism (vice versa) in analysing the *Consensus Economics* data. This result takes us a step further, showing that in the Bank of England survey forecasters with relatively low point forecasts of inflation tend to have relatively high forecasts of GDP growth, and vice versa, persistently so over this period. However, there is no association between forecasters' point forecast optimism/pessimism and their relative uncertainty described in the preceding section, with only small correlations, of mixed signs, between the two rankings, in the six cases (two variables, three horizons).



Figure 6. Point forecasts of the five highest-ranked (circles) and lowest-ranked (pluses) regular respondents in each panel. Left column: CPI inflation; right column: GDP growth; $h = 5, 9, 13$

5. UNCERTAINTY AND DISAGREEMENT REVISITED

Disagreement among several available point forecasts is often used as an indicator of uncertainty whenever a direct measure is not available. Once direct measures of uncertainty based on density forecasts become available, the utility of such proxy variables can be assessed, and a research literature has developed, mostly based on US SPF data, originating with the seminal article by Zarnowitz and Lambros (1987). We studied this question using the Bank of England survey, 1996–2005, in a previous article (Boero *et al.*, 2008), and we return to it with the present dataset, which covers a less quiescent period, and with the improved measure of uncertainty developed above.

Table III. Measures of agreement over time between forecasters' rankings with respect to their point forecasts: Kendall coefficients of concordance

	$h=5$	$h=9$	$h=13$
CPI inflation	0.20	0.27	0.15
GDP growth	0.33	0.46	0.49

An appropriate statistical model for a combined density forecast, such as the survey average forecast, is the finite mixture distribution (Wallis, 2005), and the standard expression for its variance provides useful discrimination between measures of uncertainty and disagreement. Denoting n individual density forecasts of a random variable Y at some future time as $f_i(y)$, with mean μ_i and variance σ_i^2 , $i=1, \dots, n$, the survey average density forecast is

$$f_A(y) = \frac{1}{n} \sum_{i=1}^n f_i(y)$$

with mean μ_A and variance σ_A^2 given as

$$\mu_A = \frac{1}{n} \sum_{i=1}^n \mu_i, \sigma_A^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2 + \frac{1}{n} \sum_{i=1}^n (\mu_i - \mu_A)^2$$

These expressions hold irrespective of the forms of $f_i(y)$, which might include histograms, constructed from the surveys. The last equation says that the variance of the survey average density forecast is equal to the average individual uncertainty (variance) plus a measure of the dispersion of, or disagreement between, the individual density forecast means. Taking these to be the relevant point forecasts, the equation shows that the disagreement between them is a component of the variance of the survey average density forecast; hence a 'pure' measure of uncertainty is the average individual variance. The research question then concerns the correlation, if any, between these two components: is disagreement a good indicator of average individual uncertainty, for use in the absence of a direct measure?

To mimic the practical situation in which only point forecasts are available, we base our disagreement measure on the reported point forecasts from the survey, which may not coincide with the density forecast means, so the equation may not hold exactly. A second practical feature is that it is usually preferred to report standard deviations, not variances, in order to show measures whose units coincide with the units of the variable under consideration, whereas the equation holds for variances, and not for standard deviations. It nevertheless provides a useful conceptual framework for the problem, in particular making clear that the dispersion of the survey average density forecast is not an appropriate indicator of forecasters' uncertainty, because it includes a component that approximates disagreement between their point forecasts.

Our practical measures corresponding to the three terms in the equation are then:

- (i) the standard deviation of the survey average density forecast, estimated via a fitted normal distribution;
- (ii) the square root of the average of the individual variances estimated as above, termed the root mean subjective variance (RMSV) in our 2008 article, after Batchelor and Dua (1996);
- (iii) a robust quasi-standard deviation (qsd) measure of disagreement, also used in our 2008 article, following Giordani and Soderlind (2003), given as one-half of the difference between the 16th

and 84th percentiles of the sample of point forecasts, appropriately interpolated. (For a normal distribution, this interval is equal to the mean \pm 1 standard deviation.)

These three measures are plotted in the six panels of Figure 7; note that, as in several preceding figures, different scales are used for the inflation and GDP growth panels.

In all six panels of Figure 7 the experience over the first 2 years shown is very similar to that seen over the earlier years of this decade in our previous article, with low disagreement and relatively little



Figure 7. Aggregate variation, average individual uncertainty and disagreement. Left column: CPI inflation; right column: GDP growth; $h = 5, 9, 13$

movement in the series, and hence little possibility of interesting co-movements. The picture then changes dramatically with the onset of the crisis, as suggested in some of the preceding figures. There are rapid increases in all three series in all six panels, most prominently so in the disagreement measure, as anticipated in Figure 5. These pronounced movements in common result in the high correlations shown in Table IV, and the conclusion that, over this period, changes in disagreement are associated with changes in uncertainty. The changes in uncertainty are proportionately smaller in all cases; in the four cases with prominent spikes in disagreement they contribute less to changes in the variance of the survey average density forecast than the changes in the disagreement measure.

The switch from a negative to a positive answer to the question ‘Is disagreement a useful proxy for uncertainty?’ between our earlier article and the present work is a mirror image of research findings on the US SPF data. Zarnowitz and Lambros originally found ‘some direct empirical support ... that greater interpersonal differentiation of expectations is a symptom of greater uncertainty’ (Zarnowitz and Lambros 1987, p. 607), using data from the survey’s start, in late 1968, to 1981. In contrast, Lahiri and Liu (2006) and Rich and Tracy (2010) give negative answers, both articles being based on a much longer sample, from 1968 to the early 2000s. The period from 1968 to 1981 is dominated by the Great Inflation, while the longer period adds in the Great Moderation; a significant reduction in the volatility of US inflation and output in the early 1980s has been widely documented. Thus the joint results from the US and UK surveys suggest the encompassing conclusion that disagreement is a useful proxy for uncertainty when it exhibits large fluctuations, but low-level high-frequency variations are not sufficiently correlated. Updating of the US studies to the recent crisis period is awaited.

6. CONCLUSION

In this article we consider several statistical issues that arise in the construction and interpretation of measures of forecast uncertainty from individual density forecasts obtained by surveying forecasters. We find substantial heterogeneity in forecasters’ uncertainty about future outcomes, as expressed in their subjective probabilities, and strong persistence in the relative level of individual forecasters’ uncertainty. This is a new finding in the context of surveys of professional forecasters, a popular research vehicle. It demonstrates individual characteristics of forecaster responses that merit deeper investigation. Using the same statistical procedures we also find persistence, at a lower level, in individual forecasters’ relative point forecasts, reflecting their relative optimism or pessimism about future prospects for inflation and GDP growth. This is not a new finding for these variables taken separately, but we also establish that a bivariate relation exists, jointly defining a persistently optimistic forecast as one of relatively low inflation and high growth, and pessimism vice versa.

Our experience in conducting this research also leads to suggestions for improving the reporting of survey results. Since the available samples of these demonstrably heterogeneous forecasters are not large, and vary over time, comparison of summary results between the current and preceding surveys, which is a common practice, can be strongly affected by missing observations in one or other survey. Our recommendation is that such comparisons be based only on the individual respondents who are present in both surveys. Secondly, although information about disagreement is often supplied, as a

Table IV. Correlation coefficients between uncertainty (RMSV) and disagreement (qsd)

	$h=5$	$h=9$	$h=13$
CPI inflation	0.77	0.76	0.62
GDP growth	0.54	0.67	0.81

histogram of point forecasts, for example (also an example of a 'this quarter/last quarter' comparison just mentioned), little is typically reported about uncertainty. A table of survey average probabilities is standard, but often its only use is to describe the survey average probability that future inflation or growth will lie to one side or the other of a threshold of interest. However, a measure of average individual uncertainty could be derived, without replicating all our calculations described above, as the difference between an estimate of the variance of the survey average density forecast and the variance of the histogram of point forecasts (disagreement): the variance equation in Section 5 above then yields the implied mean subjective variance.

ACKNOWLEDGEMENTS

An earlier version of this article was presented at the Seventh ECB Workshop on Forecasting Techniques, Frankfurt, May 2012, and the 32nd International Symposium on Forecasting, Boston, June 2012. We are grateful to discussants and referees for helpful comments and suggestions, and to Bank of England staff for assembling 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.

REFERENCES

- Batchelor RA. 2007. Bias in macroeconomic forecasts. *International Journal of Forecasting* **23**: 189–203.
- Batchelor RA, Dua P. 1996. Empirical measures of inflation uncertainty: a cautionary note. *Applied Economics* **28**: 333–341.
- Boero G, Smith J, Wallis KF. 2008. Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *Economic Journal* **118**: 1107–1127.
- Bruine de Bruin W, Manski CF, Topa G, van der Klaauw W. 2011. Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics* **26**: 454–478.
- D'Amico S, Orphanides A. 2008. Uncertainty and disagreement in economic forecasting. Finance and Economics Discussion Series 2008-56. Federal Reserve Board, Washington DC.
- Engelberg J, Manski CF, Williams J. 2009. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics* **27**: 30–41.
- Giordani P, Soderlind P. 2003. Inflation forecast uncertainty. *European Economic Review* **47**: 1037–1059.
- Johnson NL, Kotz S, Balakrishnan N. 1995. Continuous Univariate Distributions, Vol. 2 (2nd edn). Wiley: New York.
- Kendall MG, Gibbons JD. 1990. Rank Correlation Methods (5th edn). Edward Arnold: London.
- King MA. 2003. Speech to the East Midlands Development Agency, 14 October 2003. Available: <http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2003/speech204.pdf>
- Lahiri K, Liu F. 2006. Modelling multi-period inflation uncertainty using a panel of density forecasts. *Journal of Applied Econometrics* **21**: 1199–1219.
- Manski CF, Molinari F. 2010. Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics* **28**: 219–231.
- Patton AJ, Timmermann A. 2010. Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics* **57**: 803–820.
- Rich R, Tracy J. 2010. The relationships among expected inflation, disagreement, and uncertainty: evidence from matched point and density forecasts. *Review of Economics and Statistics* **92**: 200–207.
- Soderlind P. 2011. Inflation risk premia and survey evidence on macroeconomic uncertainty. *International Journal of Central Banking* **7**: 113–133.
- Wallis KF. 2005. Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics* **67**: 983–994.
- Zarnowitz V, Lambros LA. 1987. Consensus and uncertainty in economic prediction. *Journal of Political Economy* **95**: 591–621.

APPENDIX: FITTING NORMAL DISTRIBUTIONS TO TWO-BIN HISTOGRAMS

We assert on page 3, in discussing the beta distribution, that ‘As is the case with the normal distribution, non-zero probabilities are needed in at least three bins to allow fitting to proceed’. On the contrary, the claim that a normal distribution can be fitted by least squares to a two-bin histogram is made by Lahiri and Liu (2006, p. 1204) and Soderlind (2011, p. 119), with reference to the software used by Giordani and Soderlind (2003). However, this is incorrect. The best-fitting distribution in such cases is a degenerate distribution with zero variance, whereas the estimate returned by the software is determined by the stopping rule employed in the numerical optimization routine. There is no check that the reported variance estimate corresponds to a local minimum of the error function, and on restarting the program with a tighter convergence criterion a reduced estimate is obtained. In discussing the two-bin case, D’Amico and Orphanides (2008, p. 9) state that ‘estimation mechanically identifies the best fitting normal distribution to be the one with infinitesimal (zero) variance and a mean located at the point dividing the two consecutive bins’, which is a correct description of the ultimate solution, but incorrect insofar as the software does not achieve it.

An example in which the density forecast of inflation assigns non-zero probabilities only to the two central bins, 1.5–2% and 2–2.5%, respectively 0.3 and 0.7, is shown in Figure A1. The forecast cumulative distribution function (CDF) $F(x)$ is zero until $x = 1.5$, then $F(2) = 0.3$, finally $F(2.5) = 1$, where it stays for $x > 2.5$. The question is what could be assumed about the distribution of probability within the bins. The conventional ‘uniform within bins’ assumption joins the points (1.5, 0), (2, 0.3), (2.5, 1) with straight-line segments. The normal approximation is steeper in the centre of the range and less steep in the tails, as desired. However, when fitting a normal CDF to these points, the error is monotonic in σ , approaching zero as $\sigma \rightarrow 0$; hence the best-fitting normal distribution is the degenerate case with $\sigma = 0$, whose CDF is the step function shown in the figure. How closely the software’s ‘least squares’ estimate approximates the step function is determined by the chosen stopping rule, and ultimately by the limit of machine accuracy. The reported variance, which is arbitrarily close to zero, is not a useful indicator of the implied forecast uncertainty.

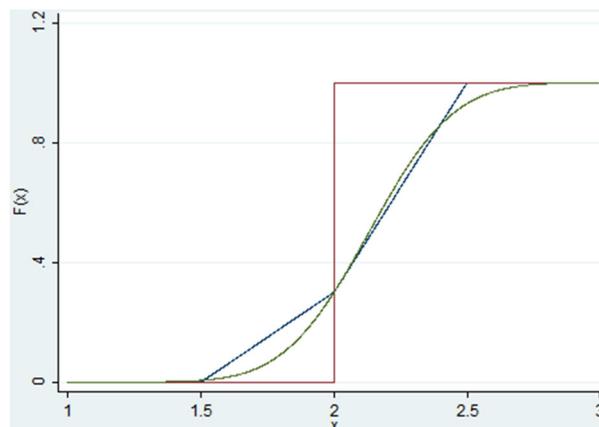


Figure A1. Cumulative distribution functions for a two-bin (30%, 70%) histogram. Piecewise linear: uniform-within-bins; continuous curve: a normal approximation; step function: the limiting degenerate case