

Internal consistency of survey respondents' forecasts:
Evidence based on the Survey of Professional Forecasters

No 772

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters.

Michael P. Clements*

Department of Economics,

University of Warwick,

UK. CV4 7AL.

m.p.clements@warwick.ac.uk.

October 16, 2006

Abstract

We ask whether the different types of forecasts made by individual survey respondents are mutually consistent, using the SPF survey data. We compare the point forecasts and central tendencies of probability distributions matched by individual respondent, and compare the forecast probabilities of declines in output with the probabilities implied by the probability distributions. When the expected associations between these different types of forecasts do not hold for some individuals, we consider whether the discrepancies we observe are consistent with rational behaviour by agents with asymmetric loss functions.

Journal of Economic Literature classification: C53, E32, E37

Keywords: Rationality, probability forecasts, probability distributions.

1 Introduction

There is a large literature addressing the rationality, efficiency and accuracy of various types of forecasts. In the economics sphere point forecasts of the conditional mean or some other measure of the central tendency have been the main focus (see, e.g., Mincer and Zarnowitz (1969), Figlewski and Wachtel (1981), Zarnowitz (1985), Keane and Runkle (1990), Davies and Lahiri (1995), Stekler (2002) and Fildes and Stekler (2002)),

*Computations were performed using code written in the Gauss Programming Language. Paul Söderlind kindly made available the code used in Giordani and Söderlind (2003).

but there has also been research on the evaluation of probability distributions (e.g., Diebold, Gunther and Tay (1998), Diebold, Hahn and Tay (1999a), Berkowitz (2001), Thompson (2002) and Hong (2001)), interval forecasts (Granger, White and Kamstra (1989), Baillie and Bollerslev (1992), McNees (1995) and Christoffersen (1998)), volatility forecasts (Anderson and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2003)), and event and probability forecasts (e.g., Granger and Pesaran (2000a, 2000b) and Pesaran and Skouras (2002), drawing on the meteorological literature: see Dawid (1986) for a review).

All of these papers evaluate series of forecasts of the same type, in terms of testing for biasedness, or efficiency or some other notion of rationality. Originally, attention focused on series of forecasts of a fixed horizon made at different points of time by a single agent. More recently, approaches have been developed to allow the analysis of point forecasts made by multiple individuals, of a number of lead times, and made at a number of different forecast origins (Davies and Lahiri (1995, 1999)). This literature has addressed the validity of various expectations formation mechanisms, such as rational expectations.

The novel aspect of our paper is that we consider forecaster rationality in terms of the internal consistency of the different types of forecasts simultaneously made by individual forecasters. This is an area which appears to have gone largely unresearched, but offers an alternative angle on the notion of individual forecaster rationality, and complements the standard approaches in the literature. Our source of forecasts is the SPF¹ survey data. This is a unique resource, in that it allows us to derive matched point, probability and probability distribution forecasts for individual forecasters. Put simply, suppose a forecaster F reports point forecasts x , probability forecasts p and probability distributions f_X , all related to the same phenomenon, then intra-forecaster consistency places restrictions on the values that these forecasts would be expected to take. In this paper we look for a number of expected associations between the different types of forecasts. These may fail for any individual due to reporting errors or other idiosyncratic factors, but if forecasters taken as a group are internally consistent, then we would expect these relationships to hold for a majority of the individual respondents. We begin by analysing whether these expected relationships do hold by taking all individual forecasters together, but considering forecasts of different horizons separately. Our main focus is on whether forecasters' point forecasts are consistent with the central tendencies of their reported probability distributions, and whether their reported probability forecasts of a decline in output are consistent with their probability distributions. Other comparisons are possible, but require additional subsidiary assumptions about forecaster behaviour that may not hold. These are noted.

¹This is a quarterly survey of macroeconomic forecasters of the US economy that began in 1968, administered by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER). Since June 1990 it has been run by the Philadelphia Fed, renamed as the Survey of Professional Forecasters (SPF): see Zarnowitz (1969) and Croushore (1993).

Because the SPF respondents report their probability distributions as histograms, a difficulty is inferring the measures of central tendency from the histograms. To compare the point forecasts and the histograms, we follow Engelberg, Manski and Williams (2006) in calculating bounds on the moments. We then adapt their idea in a simple way to calculate bounds on the probabilities of decline from the histograms. The point and probability forecasts are then compared to the relevant bounds. Focusing on the conditional mean, our findings for point forecasts match those of Engelberg *et al.* (2006), in that most, but not all, point forecasts lie within the bounds on the histogram means, and violations that exceed the upper bound are more common than those in the other direction. Our results for the probability forecasts show a far greater proportion of violations even though the bounds are often wide (in a sense explained below). Nevertheless, the probability forecasts follow the point forecasts in that they also tend to be favourable, in the sense that the forecast probabilities of decline tend to understate the probabilities derived from the histograms.

We then consider a leading explanation for these findings: that forecasters' loss functions are asymmetric. If this is the case, rationality does not pre-empt the forecaster reporting as a point forecast a quantile of their probability distribution that is arbitrarily far from the central tendency for a high enough degree of asymmetry. As individuals may have different loss functions with different degrees of asymmetry, this line of investigation leads to a consideration of forecasts at an individual level, even though the nature of the SPF data set is such that the number of observations for a specific individual may be relatively small. We consider an approach to testing for asymmetry that does not require rationality, in the sense of an efficient use of all available information. We show that the usual recommendation for testing for asymmetry requires rationality, in this sense. As our test is based on OLS, it may be more robust for small data samples than methods such as Elliott, Komunjer and Timmermann (2005b) based on GMM, although the informational requirements of our method will limit its general applicability.

Care is required in ensuring that the different types of forecasts do relate to the same object, in terms of horizon and timing of the forecast. In the next section we discuss the nature of the different types of forecasts reported in the SPF, and the ways in which we attempt to ensure a fair comparison. The SPF point forecasts and histograms have been widely analysed in separate exercises.² The probability forecasts (of the event that output will decline) have received relatively little attention.³ A seminal paper by Zarnowitz and

²For example, the point forecasts have been analysed by Zarnowitz (1985), Keane and Runkle (1990), Davies and Lahiri (1999), and the probability distributions by Diebold, Tay and Wallis (1999b), Giordani and Söderlind (2003) and Clements (2006). A detailed academic bibliography of papers that use SPF data is maintained at <http://www.phil.frb.org/econ/spf/spfbib.html>.

³Although the SPF produces an 'anxious index' by averaging the individual respondents' probabilities of declines in real output in the following quarter, and this is shown to be correlated with the NBER business cycle periods of expansion and recession.

Lambros (1987)⁴ drew on different types of forecasts, namely the individual respondents' point forecasts and their expected probability distributions for inflation and output, to see whether measures of dispersion or 'disagreement' across forecasters could be used as a proxy for aggregate uncertainty (as measured by the average variance or standard deviation of the individuals' histograms). Our contribution is quite different as it involves the comparison of these (and other forecasts) for a given individual.

The plan of the rest of the paper is as follows. Section 2 discusses the SPF, from which we obtain the forecast data. Section 3 reports the comparison of the histograms and point forecasts across all forecasters using the Engelberg *et al.* (2006) non-parametric bounds approach. Section 4 reports an extension of this approach to the comparison of the histograms and probabilities of decline. We also consider whether favourable point and probability forecasts tend to be issued simultaneously by an individual. Section 5 reports the individual-level analysis of the hypothesis that forecasters have asymmetric loss functions, as a possible explanation of the findings that individuals' point forecasts are sometimes too favourable relative to their histograms. In section 6 we take a stand on what it is that the forecasters are trying to forecast - whether they are forecasting an early announcement of the data or a later revision. Up to this point, we have not needed to make any assumptions about the actual values of the series against which a forecaster would wish their forecast to be judged, as forecasts are not compared to outcomes at any stage. By making such an assumption we are able to assess the asymmetric loss hypothesis in a way that complements the analysis of section 5.

Section 7 considers other comparisons that could be drawn between the different types of forecasts and the assumptions that need to be made to make these feasible. Section 8 provides some concluding remarks.

2 The Survey of Professional Forecasters (SPF)

The SPF quarterly survey began as the NBER-ASA survey in 1968:4 and runs to the present day. The survey questions elicit information from the respondents on their point forecasts for a number of variables; their histograms for output growth and inflation; and the probabilities they attach to declines in real output. The surveys are sent in the first month of the quarter, so the respondents will know the advance GDP release for the previous quarter when they make their forecasts. The Federal Reserve Bank of Philadelphia maintains a quarterly Real Time Data Set for Macroeconomists (RTDSM).⁵ This consists of a data set for each quarter that contains only those data that would have been available at a given reference date: subsequent revisions,

⁴For recent contributions, see Rich and Tracy (2003) and Giordani and Söderlind (2003).

⁵See Croushore and Stark (2001).

base-year and other definitional changes that occurred after the reference date are omitted.⁶ The RTDSMs tie in with the SPF surveys such that a respondent to, say, the 1995:Q1 survey would have access to the data in the Feb 1995 RTDSM. The datasets contain quarterly observations over the period 1947:Q1 to the quarter before the reference date (1994:4, in our example). Our focus is on the forecasts of real output, as for this variable alone we have point and probability forecasts (of decline) as well as histograms. We use data from 1981:3 to 2005:1, as prior to 1981:3 the histograms for output growth referred to nominal output, and point forecasts for real GDP (GNP) were not recorded.⁷ The definition of real output changes from GNP to GDP in 1992:1, and there are also base year changes over the period. Base year changes may potentially be important, but are taken care of by using the RTDSMs.

Over our sample period (the 1981:3 to 2005:1 surveys), respondents are asked to give their point forecasts for the level of output in the current year, as well as for the previous quarter, for the current quarter, the next four quarters, and for the current year.⁸ For the probabilities of declines in real GDP (real GNP prior to 1992), respondents are asked to give probabilities of a decline in the current quarter (the survey date quarter) relative to the previous quarter, for the next quarter relative to the current quarter, and so on up to the same quarter a year ahead relative to three quarters ahead.⁹ The probability distributions refer to the annual change from the previous year to the year of the survey, as well as of the survey year to the following year, and we use only the former. Therefore, the Q1 surveys provide four-quarter ahead probability distribution forecasts (as histograms) of the annual change in the current year over the previous year, which can be matched to the annual output growth point forecasts (obtained from the annual level forecasts).¹⁰

⁶Later vintages of data would not serve our purpose, as they typically contain revisions and definitional changes that were largely unpredictable (see Faust, Rogers and Wright (2005)) based on information available at the time, and as such would not have been known to the SPF respondents. Croushore and Stark (2001, 2003), Koenig, Dolmas and Piger (2003), Patterson (1995, 2003) and Faust, Rogers and Wright (2003) discuss issues related to data vintages.

⁷Engelberg *et al.* (2006) further restrict their start period to 1992:1, in part because it is not clear that the respondents were always provided with the previous year's level of output prior to this period, so that when we calculate the annual growth point forecast it is not clear whether the base value matches that used by the forecaster. As some of our empirical analysis requires the construction of time-series of forecasts for individual forecasters, we can not afford to lose the data for the 1980s. As the value in the previous year would have been in the public domain at the time, and is now recorded in the RTDSMs, it seems unlikely that constructing annual output growth forecasts based on the values and vintages that were in the public domain at the time would be too distortionary.

⁸The most recent figures are sent out with the survey forms, so that a respondent to the 95:Q1 survey would receive the values for the variables for which point forecasts are required up to and including 94:Q4. Most respondents use the 94:Q4 figures as their forecasts of that quarter.

⁹So again using the example of the 1995 Q1 survey to fix ideas, respondents are asked to report the probabilities they attach to declines in 95:Q1 relative to 94:Q4, 95:Q2 relative to 95:Q1, 95:Q3 relative to 95:Q2, 95:Q4 relative to 95:Q3 and 96:Q1 relative to 95:Q4.

¹⁰The point forecasts of the growth rate are calculated using the actual data for the previous year from the RTDSM available in the quarter of the survey. The one exception is that the RTDSM for 1996Q1 is missing the value for 1995Q4. In constructing the year-on-year point forecast growth rates for the respondents to the 1996Q1 survey we use the previous-quarter forecasts (of 1995Q4).

Prior to 1981:3, annual point forecasts were not reported, so authors such as Zarnowitz and Lambros (1987), using the earlier surveys, calculated annual forecasts from the forecasts of the quarters. Engelberg *et al.* (2006) note that summing the forecasts of the quarters to obtain the annual forecast is only valid if the point forecasts are all means; otherwise the adding up property does not hold.

The Q2 surveys provide three-quarter ahead forecasts, down to one-quarter ahead forecasts from the Q4 surveys. Thus we obtain a matched pair of forecasts of a single horizon from each survey.

When we consider the comparison of the histograms and probabilities of decline in section 4 we will show that only the Q4 survey data can be used, and in section 7 discuss how the quarterly point forecasts and decline probability forecasts might be compared to give a much larger number of observations.

The total number of usable point and probability forecasts across all surveys and respondents is 2462. These forecasts come from the 95 quarterly surveys from 1981:3 to 2005:1, and from 181 different respondents. We restrict the sample to only include regular forecasters - those who have responded to more than 12 surveys.¹¹ This gives 73 respondents. These regular respondents account for 1969 forecasts, some 80% of the total.

3 Consistency of the point forecasts and probability distributions

It is unclear whether the point forecasts should be interpreted as the means, modes or even medians of the probability distributions. Engelberg *et al.* (2006) calculate non-parametric bounds on these measures of central tendency for the histograms. We focus on the mean, as their results are similar across the three measures. Before calculating bounds, we briefly consider other approaches that have been adopted to calculate means from histograms. One approach in the literature is to assume that the probability mass is uniform within a bin (see e.g., Diebold *et al.* (1999b), who make this assumption in the context of calculating probability integral transforms). Another is to fit normal distributions to the histograms (see Giordani and Söderlind (2003, p. 1044)). If the distribution underlying the histogram is approximately ‘bell-shaped’ then the uniformity assumption will tend to overstate the dispersion of the distribution because there will be more mass close to the mean. This problem will be accentuated when there is a large difference in the probability mass attached to adjacent bins, where it might be thought desirable to attach higher probabilities to points near the boundary with the high probability bin. In the same spirit of fitting a parametric distribution to the histogram, Engelberg *et al.* (2006) argue in favour of the unimodal generalized beta distribution. Using the assumption of uniform mass within a bin, and approximating the histogram by a normal density, results in the correlations between the annual growth point forecasts and the histogram means recorded in table 1.

The crude correlations in columns 3 and 4 are all high, between 0.92 and 0.96, and the results indicate little difference between the two methods of calculating histogram means that we consider. Nevertheless, it is

¹¹One might suppose that regular respondents are more au fait with the survey and what is required of them, so that errors in filling in the survey questionnaires are less likely.

Table 1: The point forecasts and the means of the probability distributions (histograms).

Survey quarter	# of forecasts	Hist. ¹	Hist. ⁿ	% within mean bounds	% below mean bounds	% above mean bounds
Q1	451.000	0.920	0.927	74.723	4.878	20.399
Q2	487.000	0.941	0.946	80.903	3.696	15.400
Q3	439.000	0.922	0.925	81.777	3.645	14.579
Q4	408.000	0.953	0.958	89.706	3.431	6.863

We exclude the Q1 surveys of 1985 and 1986 as the Philadelphia Fed has documented problems with the forecast distributions made in the first quarters of 1985 and 1986 – there is some ambiguity about the periods to which the survey returns apply.

Hist.¹ denotes the means are calculated directly from the histograms. Hist.ⁿ denotes the means are calculated from the histogram using a normal approximation to the histogram.

difficult to infer much about the consistency of the probability distributions and the point forecasts from these correlations, especially as they rely upon the assumptions we have made to calculate the histogram means. We can dispense with these assumptions if we adopt the bounds approach of Engelberg *et al.* (2006).¹² The last three columns of the table present the percentage of point forecasts that lie within, below, and above, the bounds respectively. The Q1 figure of 74.7 indicates that about three-quarters of all first-quarter survey point forecasts of annual growth lie within the bounds on the histogram mean. Alternatively, one quarter of point forecasts are not consistent with the histogram means. Of this quarter, approximately 80% exceed the upper limit, indicating a preponderance of ‘favourable’ point forecasts. We find that the proportion which are consistent increases with the survey quarter, corresponding to a shortening horizon, although the tendency to report favourable point forecasts persists. These results are similar to those of Engelberg *et al.* (2006), except that we find higher proportions of forecasts inconsistent with the mean, especially for Q1. As noted in the introduction, respondents may rationally report any quantile of their probability distributions if they have asymmetric loss.

4 Consistency of the probability forecasts and probability distributions

The nature of the SPF histograms and probability forecasts is such that a comparison of the two can only be made for the Q4 surveys. For these surveys we are able to calculate current-quarter probabilities of decline

¹²The lower (upper) bound is calculated by assuming that all the probability lies at the lower (upper) limit of the histogram bin or interval.

from the probability distributions, which can be compared to the directly-reported current-quarter forecast probabilities of decline. We are unable to infer the probability of decline in a specific quarter for surveys other than the Q4 surveys - the Q4 survey histograms relate to annual growth in the current year, but the only unknown is the value of output in the current quarter (the fourth quarter). Given the realized values of output in the seven quarters to the fourth quarter (taken from the appropriate RTDSM), we can infer the year-on-year rate of growth that equates the Q4 level of output with that of the preceding quarter. The implied current-quarter probability of decline from the histogram is then the probability that year-on-year output growth will not exceed this rate. As in the case of calculating means from the histograms, the required calculation could be performed by assuming uniformity within intervals, or by approximating the histograms by normal densities, amongst other methods. But as in the analysis of the point forecasts and histogram means, rather than considering the correlation between the forecast probability series and the derived series, calculating bounds allows us to compare the histograms and probability forecasts without making an assumption about the relationship between the histograms and the respondents' actual beliefs.

To see how the bounds are calculated, consider the following example. Suppose the forecaster attaches probabilities of 0.1, 0.5 and 0.4 to the intervals $[3, 4)$, $[4, 5)$ and $[5, 6)$, with all other bins having zero probabilities. Suppose output will decline if $y < 4.2$. An upper bound on the probability (when all mass in the $[4, 5)$ interval is less than 4.2) is $0.1 + 0.5 = 0.6$, and the lower bound is 0.1, when all mass in the $[4, 5)$ interval is greater than 4.2. Thus the lower bound sums the probabilities of all intervals whose upper bounds are less than the value (here 4.2), and the upper bound includes in this sum the probability attached to the interval containing the value. So the size of the bound is the probability attached to the interval containing the value, and may as a consequence be large. At the extremes, suppose y lies below the lowest interval. Then the upper and lower bounds on probability coincide at zero. If Y lies above the highest interval, both bounds are 1. Note we are assuming that the upper and lower histogram bins are closed, with the upper limit to the upper bin, and the lower limit to the lower bin, set so that the bin widths are the same as those of interior bins. Bounds calculated in this way satisfy $u, l \in [0, 1]$, and $u - l \in [0, 1]$, where u and l are the upper and lower bounds on the probability.

Table 2: Bounds on histogram probabilities of decline and directly-reported probabilities.

Survey quarter	# of forecasts	% within bounds	% below bounds	% above bounds
Q4	408.000	56.127	42.157	1.716

The findings in table 2 indicate that only just over a half of all probability forecasts of a decline in output in quarter 4 relative to quarter 3 are consistent with the bounds on this probability calculated from the probability distributions. As for the point forecasts, respondents as a whole report more favourable probability assessments (that is, a lower probability of a decline in output) than is suggested by their probability distributions. Here the results are more stark - virtually all forecasts outside the bounds suggest more favourable assessments. The average width of the bands is 0.69.

Also of interest is whether there is a match between the favourable point and probability forecasts: is it the case that when a favourable point forecast is issued, this tends to be matched by a favourable probability forecast? Table 3 provides this information. The first row takes all the point forecasts that fell below the lower mean bound, and reports the percentage of times the corresponding probability forecasts fell below, within and above the bounds on the probability of decline calculated from the probability distributions. Similarly for the next two rows, which condition on the point forecasts being within, and above, the bounds, respectively. The principal point to note is that over four fifths (82.4%) of favourable point forecasts (recall this is shorthand for forecasts above the upper band) are simultaneously reported with favourable (below the bound) probability forecasts, so that for the most part forecasters are consistent in reporting favourable assessments on both fronts.

Table 3: Coincidence of favourable and unfavourable point and probability forecasts.

Percentage of probability forecasts corresponding to each category of point forecast			
	below	within	above
below	5.9	58.8	35.3
within	45.5	59.0	0.5
above	82.4	17.6	0.0

To make further progress, in the next section we consider the empirical relationships between the different types of forecasts at an individual level. We have so far considered the totality of forecasts, discriminated by survey quarter (or forecast horizon) but not by individual respondent. We will consider whether there are individual level effects, such as asymmetries in individual loss functions, which can account for these findings.

5 Individual level analysis of point forecasts and histograms

Figure 1 provides an indication of whether a small number of individual respondents account for the majority of the instances when the point forecasts fall outside mean bounds. The height of each bin gives the number of respondents (out of a maximum of around 70) for whom a given percentage of forecasts fall outside the bounds (the intervals are 0-5,5-10 etc., with the midpoints plotted). We show the results for all survey quarters, and separately by quarter. For all, we find 19 respondents with 0 to 5% of their forecasts outside their mean bounds. We find 40 respondents with in excess of 10% of their forecasts outside the bands. There are differences over quarters, and in particular, a higher number (50) of individuals falling in the 0 to 5% interval when we consider fourth quarter forecasts alone. Based on these results, which show that the occurrences of bounds violations by point forecasts is not confined to a few individuals, we test all the individuals for asymmetric loss functions for whom it is practical to do so.

The strategy of only considering violations of the bounds as evidence against the hypothesis that conditional means are reported as point forecasts is necessarily a conservative strategy: a forecaster's point forecast may satisfy asymmetric loss but still fall within the mean bounds if the forecast uncertainty at that point is low and/or the degree of asymmetry is low, or if the bounds are wide. Moreover, the maintained assumption in the literature on testing for asymmetric loss is that the degree of asymmetry of the loss function is constant for a given individual over time, so that it makes sense to use all the observations for that individual. Using only those observations for which bounds violations occur would likely bias the findings in favour of asymmetry. Thus, the bounds analysis suggests that point forecasts are sometimes too favourable to be consistent with individuals reporting their conditional means, and one possible explanation is asymmetric loss. But for a fair test of the hypothesis that asymmetric loss accounts for the apparent inconsistencies between point forecasts and probability distributions, we need to consider all observations (for each individual) and not only those that violate the bounds. Further, for the statistical test described below, we need a point estimate of the individual's conditional expectation rather than a bound, so that the sharp results we obtain are at the cost of the assumption we make about the relationship between the histogram and the individual's underlying probability distribution. The test outcomes are based on calculating conditional expectations directly from the histograms.

It is now well understood that the optimal forecasts will be biased if the loss function is asymmetric (see, *inter alia*, Granger (1969), Zellner (1986), Christoffersen and Diebold (1997, 1997)). Much of the recent literature has sought to test whether forecasts are rational once we allow forecasters to have asymmetric loss functions: see, for example, Patton and Timmermann (2006), Elliott *et al.* (2005b) and Elliott, Ko-

munjer and Timmermann (2005a). Capistrán and Timmermann (2005) consider whether asymmetric loss can explain the empirical correlation that appears to hold between the dispersion of individuals' point forecasts ('disagreement') and forecast uncertainty. We consider whether asymmetric loss can account for the presentation of favourable point forecasts. The novelty of our approach to testing for rationality allowing asymmetric loss is that we do not require that individuals make use of all available information.

To motivate the approach, consider linex loss (Varian (1975)), and in order to obtain a closed-form solution for the optimal predictor, assume that the process in period $t + h$, conditional on information Ω_t available at time t , is gaussian: $y_{t+h|t} \sim N(E_t(y_{t+h}), V_t(y_{t+h}))$, where $E_t(y_{t+h}) \equiv E(y_{t+h} | \Omega_t)$, and $V_t(y_{t+h}) \equiv Var(y_{t+h} | \Omega_t)$. Linex loss is given by $L(e_{t+h,t}, \phi) = \phi^{-2} [\exp(\phi e_{t+h,t}) - \phi e_{t+h,t} - 1]$, where for $\phi > 0$, the loss function is approximately *linear* for $e_{t+h,t} < 0$ ('over-predictions'), and *exponential* for $e_{t+h,t} > 0$, ('under-predictions'). For small ϕ the loss function is approximately quadratic. It is straightforward to show that the optimal predictor satisfies $f_{t+h,t}^* = E_t(y_{t+h}) + \frac{\phi}{2} V_t(y_{t+h})$, so that the deviation between the optimal point forecast and the conditional mean depends on the conditional variance, for $|\phi| \gg 0$. A similar dependence on the conditional variance holds for other popular asymmetric loss functions. Even without the conditional normality assumption, it follows intuitively that the more costly over-predictions relative to under-predictions, say, and the greater the likelihood of over-predictions (because the more uncertain the outlook) then the more the forecaster will aim to under-predict on average. Under asymmetric loss, the bias of a rational forecaster should depend on forecast variance but should not be systematically related to other variables known at time t :

$$\begin{aligned} bias_{t+h,t} &= E(y_{t+h} - f_{t+h,t}^* | \Omega_t) \\ &= E\left(y_{t+h} - \left(E_t(y_{t+h}) + \frac{\phi}{2} V_t(y_{t+h})\right) | \Omega_t\right) \\ &= -\frac{\phi}{2} V_t(y_{t+h}) \end{aligned}$$

This motivates the suggestion of Pesaran and Weale (2006) to test for rational expectations with asymmetric losses by running a regression such as:

$$e_{t+h,t} \equiv y_{t+h} - {}_t y_{t+h} = \zeta_1 g(V_t(y_{t+h})) + \zeta_2' \mathbf{Z}_t + \epsilon_{t+h} \quad (1)$$

where under the null we would expect to find $\zeta_2' = 0$ but $\zeta_1 \neq 0$. In this regression, we have replaced the optimal forecast by the reported forecast ${}_t y_{t+h}$ to allow for reporting errors etc., but provided the two differ by an error that is a zero-mean innovation on Ω_t this switch is innocuous. Note also that $V_t(y_{t+h})$ enters

as $g(V_t(y_{t+h}))$ to allow for the possibility that for some loss functions the optimal predictor depends on the forecast standard deviation rather than variance (in which case $g(x) = \sqrt{x}$).

In the above the forecasts and information set are not indexed by the individual, and so has the interpretation that it applies to individuals who have identical loss functions and information sets. Consider now the case of heterogeneous information sets and individual-specific ϕ values, ϕ_i . Suppose forecaster i 's information set is $\Omega_{t,i} \neq \Omega_t$, then for i , $y_{t+h}|t \sim N(E_{t,i}(y_{t+h}), V_{t,i}(y_{t+h}))$, where $E_{t,i}(y_{t+h}) \equiv E(y_{t+h} | \Omega_{t,i})$, and $V_{t,i}(y_{t+h}) \equiv Var(y_{t+h} | \Omega_{t,i})$. Then $f_{t+h,t,i}^* = E_{t,i}(y_{t+h}) + \frac{\phi_i}{2} V_{t,i}(y_{t+h})$. The bias conditional on $\Omega_{t,i}$ is:

$$\begin{aligned} bias_{t+h,t,i}(|\Omega_{t,i}) &= E(y_{t+h} - f_{t+h,t,i}^* | \Omega_{t,i}) \\ &= E\left(y_{t+h} - \left(E_{t,i}(y_{t+h}) + \frac{\phi_i}{2} V_{t,i}(y_{t+h})\right) | \Omega_{t,i}\right) \\ &= -\frac{\phi_i}{2} V_{t,i}(y_{t+h}). \end{aligned}$$

We can test for efficient use of the individual's information set $\Omega_{t,i}$, assuming asymmetric loss, via a regression such as (1):

$$e_{t+h,t,i} \equiv y_{t+h} - f_{t+h,t,i}^* = \zeta_{1,i} g(V_{t,i}(y_{t+h})) + \zeta_{2,i}' \mathbf{Z}_{t,i} + \epsilon_{t+h,i} \quad (2)$$

provided the explanatory variables are restricted such that $\mathbf{Z}_{t,i} \in \Omega_{t,i}$. Suppose instead we consider the bias conditional on Ω_t . This is given by:

$$\begin{aligned} bias_{t+h,t,i}(|\Omega_t) &= E(y_{t+h} - f_{t+h,t,i}^* | \Omega_t) \\ &= E\left(y_{t+h} - \left(E_{t,i}(y_{t+h}) + \frac{\phi_i}{2} V_{t,i}(y_{t+h})\right) | \Omega_t\right) \\ &= E_t(y_{t+h} - E_{t,i}(y_{t+h})) - \frac{\phi_i}{2} E_t(V_{t,i}(y_{t+h})). \end{aligned} \quad (3)$$

Muthian rationality requires that $\Omega_{t,i}$ contains Ω_t , $\Omega_t \subseteq \Omega_{t,i}$ (and may in addition contain 'private information'), and so $E_t(E_{t,i}(y_{t+h})) = E_t(y_{t+h})$ and $E_t(V_{t,i}(y_{t+h})) = V_{t,i}(y_{t+h})$, and so has the implication that the individual's forecast bias should be correlated with their forecast variance but not with any $Z_t \in \Omega_t$. The assumption that $\Omega_t \subseteq \Omega_{t,i}$ may be stronger than we wish to make - we wish only to assume that $\Omega_{t,i}$ is used efficiently. The problem arises because $\Omega_{t,i}$ is not observed by the econometrician. Let $\xi_{t+h,i} =$

$y_{t+h} - E_{t,i}(y_{t+h})$, and suppose $E_t(\xi_{t+h,i}) \equiv E(\xi_{t+h,i} | \Omega_t) \neq 0$. Then in the regression:

$$e_{t+h,t,i} \equiv y_{t+h} - {}_t y_{t+h,i} = \zeta_{1,i} g(V_{t,i}(y_{t+h})) + \zeta'_{2,i} \mathbf{Z}_t + \epsilon_{t+h,i} \quad (4)$$

it is apparent that the disturbance term $\epsilon_{t+h,i}$ will contain $\xi_{t+h,i}$, so that the disturbance term will in general be correlated with elements of \mathbf{Z}_t which are not in $\Omega_{t,i}$, $Cov(\mathbf{Z}_t, \epsilon_{t+h,i}) \neq 0$. The OLS estimator will be inconsistent. Without knowledge of the individual's information set $\Omega_{t,i}$, it is not possible to test for a consistent use of information, or to consistently estimate $\zeta_{1,i}$ and $\zeta'_{2,i}$, without assuming $\Omega_t \subseteq \Omega_{t,i}$.

Our proposed solution is based on redefining the forecast error in the regression equation. From rearranging (3), under asymmetric loss we have that:

$$E[(y_{t+h} - f_{t+h,t,i}^* - (y_{t+h} - E_{t,i}(y_{t+h}))) | \Omega_t] = -\frac{\phi_i}{2} V_{t,i}(y_{t+h})$$

which implies that $E_{t,i}(y_{t+h}) - f_{t+h,t,i}^*$ (or $E_{t,i}(y_{t+h}) - {}_t y_{t+h,i}$) should be related to $V_{t,i}(y_{t+h})$ but should not vary systematically with any variables in Ω_t , irrespective of whether or not they are in the individual's information set. This suggests the following regression:

$$E_{t,i}(Y_{t+h}) - {}_t y_{i,t+h} = \delta_{1,i} g(V_{t,i}(Y_{t+h})) + \delta'_{2,i} \mathbf{Z}_t + \varepsilon_{t+h,i} \quad (5)$$

(for each individual $i = 1, \dots, n$, for forecasts $t = 1, \dots, T$, and for a given forecast horizon h). The null of rationality and symmetric loss for individual i is that $\delta_{1,i} = 0$, and $\delta'_{2,i} = \mathbf{0}$ against the alternative that any of these coefficients are non-zero. A rejection of the null due to $\delta_{1,i} \neq 0$, with $\delta'_{2,i} = \mathbf{0}$, indicates asymmetry (and rationality), while $\delta'_{2,i} \neq \mathbf{0}$ indicates irrationality (conditional on the assumed form of the loss function¹³). In order to carry out these tests, as well as the point forecasts we require the individual's predictive distributions ${}_t P_{i,t+h}(y) = \Pr_i(Y_{t+h} < y | \Omega_{t,i})$ so that $E_{t,i}(Y_{t+h})$ and $V_{t,i}(Y_{t+h})$ can be derived (or that these conditional moments are available directly). Estimates of these distributions are available in the SPF.

The problem with testing based on (4) can be viewed as one of measurement error, as the dependent variable in (5) is related to that in (4) by:

$$e_{t+h,i} = y_{t+h} - {}_t y_{i,t+h} = (E_{t,i}(y_{t+h}) - {}_t y_{i,t+h}) + (y_{t+h} - E_{t,i}(y_{t+h})).$$

¹³It is possible that a more complex loss function (such that loss depends on the forecast error interacted with some variable Z_t) would mean that some element of $\delta'_{2,i}$ being non-zero is not inconsistent with rationality, in the same way that asymmetric loss suggests $\delta_{1,i} \neq 0$.

We have denoted the second term on the RHS by $\xi_{t+h,i} = y_{t+h} - E_{t,i}(y_{t+h})$. In this interpretation, $\xi_{t+h,i}$ is a measurement error, such that the dependent variable in (4) is a noisy proxy for the unobserved dependent variable $E_{i,t}(Y_{t+h}) - y_{t+h,i}$ in (5). Standard analysis suggests that if $E_t(\xi_{t+h,i}) = 0$, so that $\xi_{t+h,i}$ is uncorrelated with any variables that might be included as explanatory variables, then inference based on (4) is less precise, but tests of both $\zeta_{1,i} = 0$ and $\zeta'_{2,i} = 0$ remain valid. But when $Cov(\xi_{t+h,i}, \mathbf{X}_{ti}) \neq \mathbf{0}$, where $\mathbf{X}_{ti} = [V_{t,i}(y_{t+h}) \quad \mathbf{Z}'_t]'$, then inference based on (4) is invalid.¹⁴

We estimate a regression such as (5) for each individual respondent and for each forecast horizon. As there are not many observations per respondent by forecast horizon, we set the lower limit to ten observations. We also restricted \mathbf{Z}_t to a constant, and in the results reported set $g(x) = x$. For each regression, we recorded the following information: whether the null that $\delta_{2,i} = 0$ was rejected at the 10% level when the variance term was omitted from the regression; whether the null that $\delta_{2,i} = 0$ was rejected at the 10% level when the variance term was included; and whether we did not reject the null that $\delta_{2,i} = 0$ but did reject the null that $\delta_{1,i} = 0$, i.e., that only the forecast variance was significant. We then aggregated this information over all individuals and horizons, and over all individuals for a given horizon. This summary of our findings is recorded in table 4.

The results indicate that in roughly 40% of the cases rationality is rejected if we assume symmetric loss, and in all but two or three of these cases the estimated coefficients on the constant term ($\delta_{2,i}$) are negative (not shown), in line with the finding of the favourable reporting of point forecasts. If we test for rationality allowing for asymmetric loss (column 4), then for all quarters there is a relatively modest reduction in the number for which rationality is rejected - 6, or an additional 10% of the original number. For only 10% of the 57 regressions do we find that the variance term is significant but the constant is insignificant. The coefficient on the variance term is found to be negative when it is significant, implying greater costs to under-predicting than over-predicting. Our findings suggest that asymmetry is only a partial explanation of the propensity to report favourable point forecasts.

¹⁴Consider the simplest case where the only explanatory variable is $V_{t,i}(y_{t+h})$, and the population regression is given by:

$$E_{t,i}(Y_{t+h}) - y_{t+h,i} = \delta_{i,1}g(V_{t,i}(y_{t+h})) + \varepsilon_{t+h,i}$$

where in fact $\delta_{i,1} = 0$, so that loss is symmetric. Inference is based on this regression but with the dependent variable measured with error, so that the actual regression the investigator runs is:

$$e_{t+h,t,i} = \zeta_{1,i}g(V_{t,i}(y_{t+h})) + \epsilon_{t+h,i}$$

where $\epsilon_{t+h,i} = \varepsilon_{t+h,i} + \xi_{t+h,i}$. Then $\widehat{\zeta}_{1,i} \xrightarrow{p} \delta_{1,i} + Cov(g(V_{t,i}(y_{t+h})), \varepsilon_{t+h,i} + \xi_{t+h,i}) \times Var(g(V_{t,i}(y_{t+h})))^{-1}$. Therefore, $\widehat{\zeta}_{1,i} \xrightarrow{p} Cov(g(V_{t,i}(y_{t+h})), \xi_{t+h,i}) \times Var(g(V_{t,i}(y_{t+h})))^{-1}$. If the degree of irrationality, $\xi_{t+h,i}$, is positively (negatively) correlated with the conditional variance, then we may conclude that $\delta_{1,i}$ is positive (negative).

Table 4: Testing for rationality and asymmetric loss.

	Total # regressions	3. # rejecting rationality assuming symmetry	4. # rejecting rationality assuming asymmetry	5. # Variance term significant, constant not significant.
All qtrs	57	24	18	6
Q1	16	10	5	1
Q2	16	7	6	0
Q3	12	5	3	3
Q4	13	2	4	2

Notes. Column 3 sums the number of individual regressions for which we reject $\delta_{2,i} = 0$ at the 10% level with the variance term omitted. Column 4 reports rejections of $\delta_{2,i} = 0$ at the 10% level with the variance term included, and column 5 failures to reject the null that $\delta_{2,i} = 0$ and rejection of the null that $\delta_{1,i} = 0$. For horizons greater than one, i.e., quarters 1, 2 and 3, HAC estimates of the variance-covariance matrix of the parameter estimates were used.

6 Comparison of forecasts to outturns

In section 5 we presented evidence to suggest that asymmetric loss is not able to explain the tendency to report favourable forecasts of output. However, that evidence was based on the assumption that an individual's conditional expectation is the sum of the histogram interval midpoints weighted by the reported probabilities. In this section we will assess the evidence for the asymmetry hypothesis without making this assumption. In order to do so, we compare forecasts to outcomes, noting that the expected squared error of an individual's point forecast must exceed that of the individual's conditional expectation if the point forecast minimizes (the expected value of) an asymmetric loss function. (This follows directly from the discussion of asymmetric loss and the optimal predictor in section 5). The problem with using outcomes is that it is unclear whether forecasters seek to forecast the first announcements of the data, or the second, or some later revision. To counter the uncertainty as to what should constitute the actual values when we calculate forecast errors, we assess the sensitivity of our results using two vintages of data. We use the second release of output data, as well as the latest release (2006Q1). Nevertheless, although this gives some indication of the dependence of the results on data revisions and vintage effects, the use of actual data is remains problematic. Figure 2 shows that the latest vintage data suggests higher growth rates for much of the period up to 2000, after which growth rates have been revised down relative to the real-time series. The differences between the two series are persistent and at times substantial. Individual forecasters may be trying to forecast either of these two series, or another vintage altogether.

The results of calculating mean squared errors for the forecasts (MSFEs) are displayed in the first three columns of table 5. The MSFE is the empirical estimate of the expected squared error, calculated as the mean

Table 5: Forecast accuracy of point forecasts and histogram means

	MSFEs			% below	Using bounds		MSFE ratio ²
	Hist ¹	Hist ⁿ	Point forecasts		MSFE ratio ¹	% above	
Latest vintage data (2006Q1)							
Q1	1.977	1.985	1.724	18.9	.301	49.0	.665
Q2	1.161	1.187	1.037	21.8	.321	45.2	.485
Q3	1.038	1.078	0.878	19.1	.398	43.3	.553
Q4	0.998	0.994	0.729	24.0	.516	39.7	.522
Real-time actuals (second-release)							
Q1	1.361	1.345	1.082	13.3	.274	53.8	.646
Q2	0.581	0.582	0.429	13.5	.314	32.8	.672
Q3	0.422	0.421	0.279	7.7	.811	30.5	.931
Q4	0.405	0.372	0.164	11.0	.592	22.1	1.690

Hist.¹ denotes the means are calculated directly from the histograms. Hist.ⁿ denotes the means are calculated from the histogram using a normal approximation to the histogram. The columns headed % below and % above give the percentage of forecasts for which the outcome is below and above the bounds on the histogram mean. The MSFE ratios in the adjacent columns, MSFE ratio¹ and MSFE ratio², report the ratio of the MSFEs for the most favourable histogram mean to the point forecasts, for the forecasts corresponding to actuals that fall below and above the bounds.

of the sample of observed squared forecast errors, by quarter. The MSFEs for the conditional expectation measured as either the directly-calculated mean of the histogram (Hist¹), or the mean based on a normal approximation to the histogram (Histⁿ), are markedly higher than the MSFE for the point forecast, for both sets of outcomes.¹⁵ This is clearly at odds with the explanation of the discrepancies between the histogram means and point forecasts based on asymmetric loss. This finding against the asymmetry hypothesis may be because we have not accurately captured individual's means by either calculating the means directly from the histogram or assuming normality. Of interest is whether we still find against the asymmetry hypothesis if we calculate the conditional expectation in a way that is most favourable to the asymmetry hypothesis subject to the conditional expectation remaining within the bounds of section 3. That is, we wish to minimize the MSFE for our estimates of the conditional expectation, to see whether the evidence is consistent with the expected squared error of the conditional expectation being less than that for the point forecast. The MSFE is minimized by setting the estimate of the conditional expectation equal to the outcome, when the outcome falls within the bound; to the lower bound, when the outcome is less than the lower bound; and to the upper bound, when the outcome exceeds the upper band. As the MSFE for the conditional expectation for the set

¹⁵The MSFEs when the outcomes are taken to be the real-time series are lower than when latest vintage actuals are used, but the relative ranking across forecasts is the same.

of outcomes within the bound is zero, we report in table 5 the percentage of the total number of forecasts (by quarter) for which the actual falls below the bound, and above the bound. For these two sets of forecasts and observations, we calculate the ratio of the MSFE for the conditional expectation (equal to the lower, and upper, bounds, respectively) to the MSFE for the point forecasts. Based on this most favourable scenario for the asymmetry hypothesis, we no longer find that the point forecasts are more accurate: the aforementioned ratios are less than unity for all but the Q4 forecasts when the actuals are above the bound. As we have adopted a conservative approach, finding against the asymmetry hypothesis would have constituted strong evidence that it is false, but failure to find against it in these circumstances offers it little support. The fact that the actuals lie above the histogram bounds at least twice as often as they lie below (and four times as often for Q1 forecasts, using real-time data) is consistent with the evidence in the first three columns that favours the point forecasts on MSFE over the estimates of the conditional expectation in scenarios other than that which is most favourable to the latter.

7 Further comparisons between different types of forecasts

So far we have compared the point forecasts and the histograms in terms of their assessments of current year on previous year growth, and we have also compared the Q4 survey probability forecasts of decline with the histograms. Because the respondents to the SPF typically provide pairs of real output forecasts and probability of decline forecasts for the current quarter, and each of the next four quarters, we investigate ways in which these can be compared. Let $x_{it,h}$ denote the forecast of real output by respondent i , to the survey dated quarter t , of the level of output h -steps ahead. The probability forecasts are denoted by $p_{it,h}$. We would expect that forecast declines in output ($x_{it,h} - x_{it,h-1}$) would be positively associated with forecast probabilities of decline. Let the proportionate change be $w_{it,h}$, i.e., $w_{it,h} = (x_{it,h}/x_{it,h-1}) - 1$. Further, a respondent with a negative output growth forecast is more likely to report a higher probability of decline the smaller the uncertainty with which that forecast is held. This idea can be formalised by supposing that $w_{it,h}$ is the mean of individual i 's density forecast of output growth in period $t+h$, and that this density is a member of the scale-location family. Letting W_{t+h} denote quarterly output growth in $t+h$, we have $W_{t+h} \sim D(w_{it,h}, \sigma_{w,it,h}^2)$. If we assume that D is the gaussian density, and if $\sigma_{w,it,h}^2$ were known, the implied forecast probability of a decline in output ($W_{t+h} < 0$) would be:

$$\hat{p}_{it,h} \equiv \Phi\left(-\frac{w_{it,h}}{\sigma_{w,it,h}}\right) \quad (6)$$

where $\Phi(\cdot)$ is the cdf of the standard normal. These implied probabilities, obtained from the point forecasts, can be compared to the reported probabilities. Apart from the normality assumption, the problem with this approach is that in general the individual forecast variances, $\sigma_{w,ith}^2$, will be unknown. One possibility is to fit an ARCH or GARCH¹⁶ model to the consensus forecast errors (as in Bomberger (1996) and Rich and Butler (1998)). The consensus (i.e., average) rather than individual forecast errors are typically used as individual respondents do not file returns to each survey and may file only a small number of responses in total. The drawback is that using the consensus errors supposes that the forecast variance is the same for all respondents.

We do not need to make this assumption if we restrict attention to the Q4 surveys. For the Q4 surveys, we can obtain estimates of the $\sigma_{w,ith}$ from the respondents' histograms. In Appendix A we show how to derive $\sigma_{w,ith}$ from the histogram forecast of annual year-on-year growth.

Table 6: Correlations between probability forecasts and implied probabilities of decline, Q4 surveys only, by individual.

Forecaster id.	# of forecasts	p, \hat{p}
20	16	0.240
420	14	0.708
426	13	0.726
428	13	0.764
421	13	0.253
30	12	0.556
99	12	0.308
94	12	0.308
84	11	0.889
411	11	0.700
407	10	0.723
433	10	0.778
439	10	0.492

The results of the comparisons between the probability forecasts and the probabilities of decline, \hat{p} , are recorded in table 6. Since these comparisons only use the Q4 surveys, there are only roughly a quarter as many observations as for the other exercises. Consequently, we report correlations for individuals who have filed ten or more responses to the survey. From the table it is apparent that the correlations are no more than 0.7 for around one half of the respondents. To help interpret the relationship between p and \hat{p} , figure 3

¹⁶See Engle (1982) and Bollerslev (1986).

displays a scatter plot of the probability forecasts p (vertical axis) against \hat{p} across all respondents (including those who filed fewer than 10 forecasts). The figure depicts 417 matched forecasts. It is apparent that the majority of the points lie below the 45 degree line, so that the \hat{p} overstate the directly-reported probabilities. Typically, the probability of a decline in output is believed to be low (often set at 5, 10 or 20% etc.) but this is at odds with the individual's assessment as embodied in their point forecast (and histogram variance).

Our findings suggest that whilst there tends to be favourable reporting of both point and probability forecasts compared to bounds derived from the histograms, most individuals' probability forecasts are markedly more optimistic than would be indicated by their point forecasts, corroborating the findings of section 4.

8 Conclusions

We have investigated whether the different types of forecasts simultaneously made by individual forecasters are consistent with one another. Our main focus is on whether forecasters' point forecasts are consistent with the central tendencies of their reported histograms, and whether their reported probability forecasts of a decline are consistent with their histograms. The main difficulty we face in addressing these questions is that the reported histograms do not uniquely reveal the individual's underlying probability distributions. We begin by constructing bounds on the permissible range of values for the conditional mean from the histogram. These bounds make no assumptions about the distribution of probability mass within the histogram intervals (except that the lower and upper intervals are closed). We then apply the spirit of this bounds approach of Engelberg *et al.* (2006) to the comparison of the histograms and probabilities of decline. We find that a proportion of point forecasts of output growth are not compatible with being the conditional mean, and in the majority of these instances the point forecasts are too favourable in that they lie above the upper bound. The bounds analysis of the reported probabilities of decline and the histograms suggests a similar picture: respondents as a whole report more favourable probability assessments (that is, a lower probability of a decline in output) than is suggested by their histograms. Virtually all probability forecasts outside the bounds suggest more favourable assessments, notwithstanding that the form of the histograms gives rise to wide average bounds in probability. Moreover, we find that the reporting of favourable point forecasts and probability forecasts are correlated: for the most part forecasters are consistent in reporting favourable assessments on both fronts.

A leading explanation for the tendency to report favourable point and probability forecasts is that forecasters' loss functions are asymmetric. If this is the case, rationality does not pre-empt the forecaster reporting as a point forecast a quantile of their probability distribution other than the central tendency. We

examine this explanation at an individual level as some may weigh the costs of over-prediction more heavily than the costs of under-prediction, and vice versa. Because we have both the individuals' point forecasts and (estimates of) their conditional expectations, we can test the asymmetry hypothesis directly for each individual without needing to use outcomes. This has the advantage that we do not require that the forecasts efficiently use all available information. Consider a standard way of testing for asymmetry. This regresses the forecast error (constructed from the outcomes and point forecasts) on variables known at the time the forecast was made, and the conditional variance of the forecasts. Suppose that, relative to an individual's information set, there are a series of negative shocks to output growth over the sample period, so that the individual's point forecasts tend to be too favourable - this could be taken as evidence that the individual has asymmetric loss such that over-predictions are less costly than under-predictions. Our test considers the deviation between the point forecast and the conditional expectation, rather than the point forecast and the outcome. In the above example, the conditional expectation would control for the negative shocks - it would be higher than warranted based on an information set that includes the shocks, but under the asymmetry hypothesis the deviation between the conditional expectation and the point forecast should only depend on the conditional variance of the forecast.

Our tests of the asymmetry hypothesis require that the histograms accurately reflect the individuals true beliefs, and that our methods of calculating the point estimates of the mean are consistent with these beliefs. Conditional on these assumptions, we find that asymmetry can explain only a relatively minor part of the favourable aspect of the point forecasts. Additional evidence can be brought to bear if we are prepared to take a stance on what the outcomes are against which the forecasts should be evaluated. Under asymmetric loss the point forecasts should have higher expected squared errors than the conditional expectations. This is the case if we take the scenario most favourable to the asymmetry hypothesis, but not if we calculate the conditional expectation directly from the histograms or via a normal approximation, so the evidence is inconclusive.

A possibility that we have not addressed is that forecasters may face economic incentives not to report their true beliefs: see, e.g., Ehrbeck and Waldmann (1996), Laster, Bennett and Geoum (1999) and Ottaviani and Sorensen (2006). Forecasters may act strategically in the sense that they balance their goal of minimizing forecast errors against conflicting aims, such as convincing the market that they are well-informed, or of attracting media attention. It might be argued that the anonymity of the SPF respondents removes the incentives for the pursuance of strategic motives, although if the respondents report the same forecasts to the SPF as they make public these issues remain. Typically empirical evidence of strategic behaviour is based on an analysis of series of point forecasts and outcomes, whereas the existence of point forecasts and

estimates of conditional expectations in the SPF makes it ideal for testing the asymmetry hypothesis.

Our overall finding is that there is a tendency for individuals' point and probability forecasts to present a rosier picture of the prospects for output growth than is implicit in their probability distributions, reported as histograms. It appears that endowing the respondents with asymmetric loss functions does not explain this puzzle.

References

- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica*, **71**, 579–625.
- Anderson, T. G., and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39**, 885–905.
- Baillie, R. T., and Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, **52**, 91–113.
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, **19**, 465–474.
- Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **51**, 307–327.
- Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, **28**, 381–392.
- Capistrán, C., and Timmermann, A. (2005). Disagreement and biases in inflation forecasts. Mimeo, Banco de México and UCSD.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, **39**, 841–862.
- Christoffersen, P. F., and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, **13**, 808–817.
- Clements, M. P. (2006). Evaluating the Survey of Professional Forecasters probability distributions of expected inflation based on derived event probability forecasts. *Empirical Economics*, **31**, No. 1, 49–64.
- Croushore, D. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, **November/December**, 3–13.
- Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105(November)**, 111–130.
- Croushore, D., and Stark, T. (2003). A real-time data set for macroeconomists: Does the data vintage matter?. *The Review of Economics and Statistics*, **85(3)**, 605–617.
- Davies, A., and Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, **68**, 205–227.
- Davies, A., and Lahiri, K. (1999). Re-examining the rational expectations hypothesis using panel data on

- multi-period forecasts. In Hsiao, C., Lahiri, K., Lee, L. F., and Pesaran, M. H. (eds.), *Analysis of Panels and Limited Dependent Variable Models*, pp. 226–254: Cambridge University Press.
- Dawid, A. P. (1986). Probability forecasting. In Kotz, S., Johnson, N. L., and Read, C. B. (eds.), *Encyclopedia of Statistical Sciences*, vol. 7, pp. 210–218: John Wiley and Sons.
- Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts: With applications to financial risk management. *International Economic Review*, **39**, 863–883.
- Diebold, F. X., Hahn, J. Y., and Tay, A. S. (1999a). Multivariate density forecast evaluation and calibration in financial risk management: High frequency returns on foreign exchange. *Review of Economics and Statistics*, **81**, 661–673.
- Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999b). Evaluating density forecasts of inflation: The Survey of Professional Forecasters. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger*, pp. 76–90. Oxford: Oxford University Press.
- Ehrbeck, T., and Waldmann, R. (1996). Why are professional forecasters biased? agency versus behavioral explanations. *The Quarterly Journal of Economics*, **111**(1), 21–40.
- Elliott, G., Komunjer, I., and Timmermann, A. (2005a). Biases in macroeconomic forecasts: Irrationality or asymmetric loss. Mimeo, UCSD.
- Elliott, G., Komunjer, I., and Timmermann, A. (2005b). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, **72**, 1107–1125.
- Engelberg, J., Manski, C. F., and Williams, J. (2006). Comparing the point predictions and subjective probability distributions of professional forecasters. Mimeo, Kellogg School of Management and Department of Economics, Northwestern University.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.
- Faust, J., Rogers, J. H., and Wright, J. H. (2003). Exchange rate forecasting: The errors we’ve really made. *Journal of International Economic Review*, **60**, 35–39.
- Faust, J., Rogers, J. H., and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, **37** (3), 403–417.
- Figlewski, S., and Wachtel, P. (1981). The formation of inflationary expectations. *Review of Economics and Statistics*, **63**, 1–10.
- Fildes, R. A., and Stekler, H. O. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, **24**.

- Giordani, P., and Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, **74**, 1037–1060.
- Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operations Research Quarterly*, **20**, 199–207.
- Granger, C. W. J., and Pesaran, M. H. (2000a). A decision-based approach to forecast evaluation. In Chan, W. S., Li, W. K., and Tong, H. (eds.), *Statistics and Finance: An Interface*, pp. 261–278: London: Imperial College Press.
- Granger, C. W. J., and Pesaran, M. H. (2000b). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, **19**, 537–560.
- Granger, C. W. J., White, H., and Kamstra, M. (1989). Interval forecasting: An analysis based upon ARCH-quantile estimators. *Journal of Econometrics*, **40**, 87–96.
- Hong, Y. (2001). Evaluation of out-of-sample probability density forecasts with applications to stock prices. Manuscript, Department of Economics, Cornell University.
- Keane, M. P., and Runkle, D. E. (1990). Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review*, **80**, 714–735.
- Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, **85(3)**, 618–628.
- Laster, D., Bennett, P., and Geoum, I. S. (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics*, **114(1)**, 293–318.
- McNees, S. K. (1995). Forecast uncertainty: Can it be measured?. *Federal reserve Bank of New York Discussion Paper*.
- Mincer, J., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J. (ed.), *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.
- Ottaviani, M., and Sorensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, **81**, 441–466.
- Patterson, K. D. (1995). An integrated model of the data measurement and data generation processes with an application to consumers' expenditure. *Economic Journal*, **105**, 54–76.
- Patterson, K. D. (2003). Exploiting information in vintages of time-series data. *International Journal of Forecasting*, **19**, 177–197.
- Patton, A. J., and Timmermann, A. (2006). Testing forecast optimality under unknown loss. Mimeo, London

School of Economics and UCSD.

- Pesaran, M. H., and Skouras, S. (2002). Decision-based methods for forecast evaluation. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 241–267: Oxford: Blackwells.
- Pesaran, M. H., and Weale, M. (2006). Survey expectations. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 715–776: Elsevier, North-Holland.
- Rich, R., and Tracy, J. (2003). Modeling uncertainty: Predictive accuracy as a proxy for predictive confidence. *Federal Reserve Bank of New York Staff Reports*, **no. 161**, February.
- Rich, R. W., and Butler, J. S. (1998). Disagreement as a measure of uncertainty: A comment on Bomberger. *Journal of Money, Credit and Banking*, **30**, 411–419.
- Stekler, H. O. (2002). The rationality and efficiency of individuals' forecasts. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 222–240: Oxford: Blackwells.
- Thompson, S. B. (2002). Evaluating the goodness of fit of conditional distributions, with an application to affine term structure models. Manuscript, Harvard University.
- Varian, H. R. (1975). A Bayesian approach to real estate assessment. In Fienberg, S. E., and Zellner, A. (eds.), *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, pp. 195–208. Amsterdam: North Holland.
- Zarnowitz, V. (1969). The new ASA-NBER Survey of Forecasts by Economic Statisticians. *The American Statistician*, **23**, **No. 1**, 12–16.
- Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics*, **3**, 293–311.
- Zarnowitz, V., and Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, **95**, 591–621.
- Zellner, A. (1986). Biased predictors, rationality and the evaluation of forecasts. *Economics Letters*, **21**, 45–48.

9 Appendix A

Suppose (6), i.e., $\hat{p}_{ith} \equiv \Phi\left(-\frac{w_{ith}}{\sigma_{w,ith}}\right)$, where w_{ith} is defined in the text. Denote i 's probability distribution (derived from the histogram) of year t over year $t - 1$ growth reported in the year t , quarter 4 survey as:

$$W_t \sim N(\mu_{it}, \sigma_{it}^2)$$

where μ_{it} is the annual year-on-year point forecast made in Q4, and σ_{it}^2 is the histogram/probability distribution variance. Let $W_{t,4}$ denote growth in the fourth quarter over the third quarter (of t), then $W_{t,4} = 100(X_{t,4}/\bar{X}_{t,3} - 1)$, where $X_{t,4}$ is the year t , Q4 level of output, and $\bar{X}_{t,3}$ is the known Q3 level of output. If individual i 's point forecast of $X_{t,4}$ from the Q4 survey, namely, $x_{it,4}$, is the mean of their forecast density for $X_{t,4}$, then $W_{t,4} \sim N(w_{it}, \sigma_{w,it}^2)$, where $w_{it} = E_{it}(W_{t,4}) = 100(x_{it,4}/\bar{X}_{t,3} - 1)$, and $\sigma_{w,it}^2 = var_{it}(W_{t,4})$, where the subscripts on $E()$ and $var()$ indicate that the moment calculations are specific to i . We can obtain $\sigma_{w,it}^2$ from the histogram year-on-year variance as follows. We have $\sigma_{w,it}^2 = (100/\bar{X}_{t,3})^2 var_{it}(X_{t,4})$. Moreover, $X_{t,4} = (W_t/100 + 1) \sum_{j=1}^4 \bar{X}_{t-1,j} - \sum_{j=1}^3 \bar{X}_{t,j}$, so $var_{it}(X_{t,4}) = \left(\sum_{j=1}^4 \bar{X}_{t,j}/100\right)^2 \sigma_{it}^2$. Consequently, $\sigma_{w,it}^2 = \left(\sum_{j=1}^4 \bar{X}_{t-1,j}/\bar{X}_{t,3}\right)^2 \sigma_{it}^2$. So the probability of a decline in Q4 relative to Q3 is

$$\Pr(W_{t+1,4} < 0) = \Phi\left(-\frac{w_{it}}{\sigma_{w,it}}\right).$$

This corresponds to (6) from the notational equivalencies $w_{it} = w_{ith}$ and $\sigma_{w,it} = \sigma_{w,ith}$.

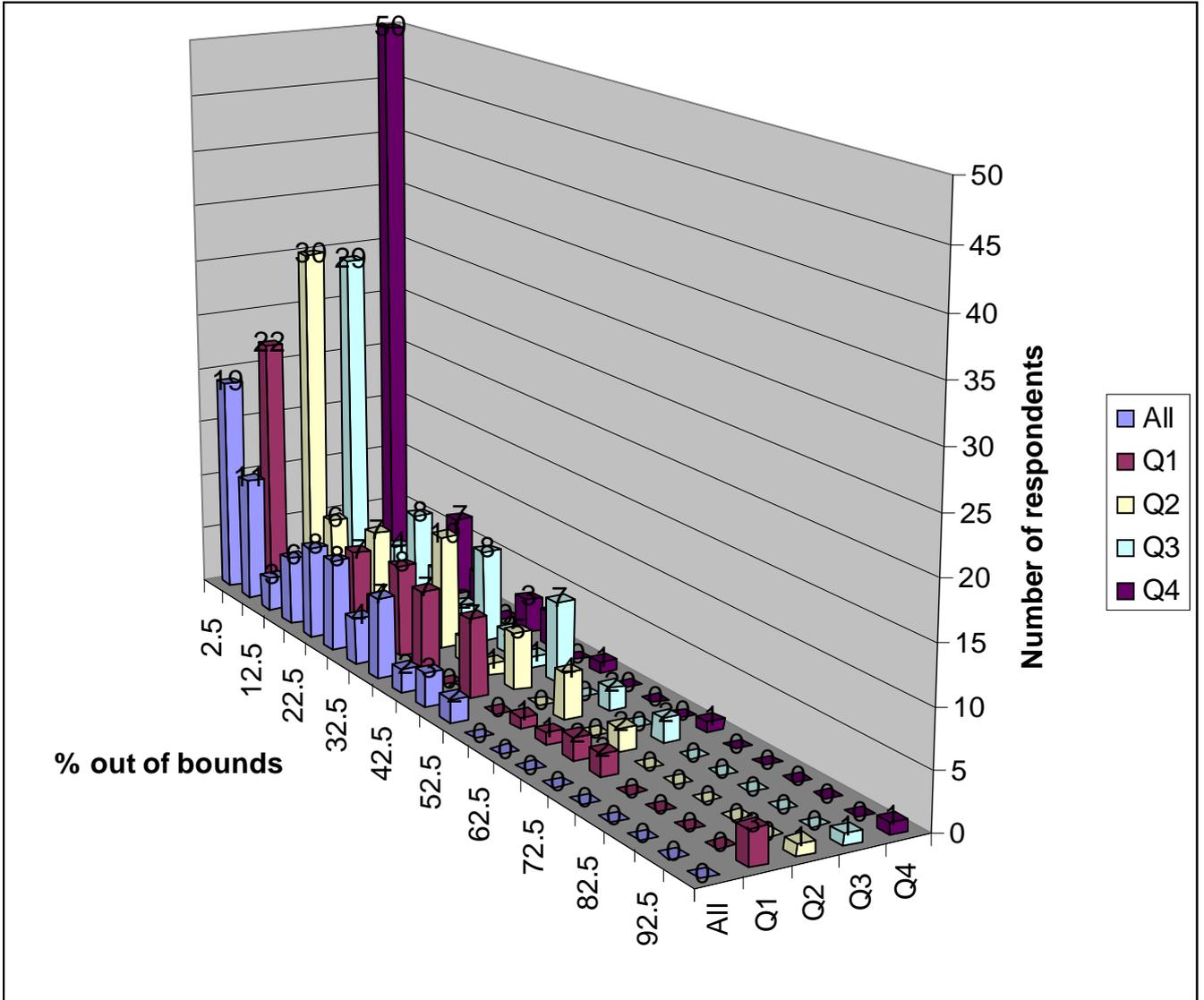


Figure 1: Distribution of individuals (count, not frequency) by percentage of point forecasts outside their mean bounds, for all quarters together, and by quarter.

growth data

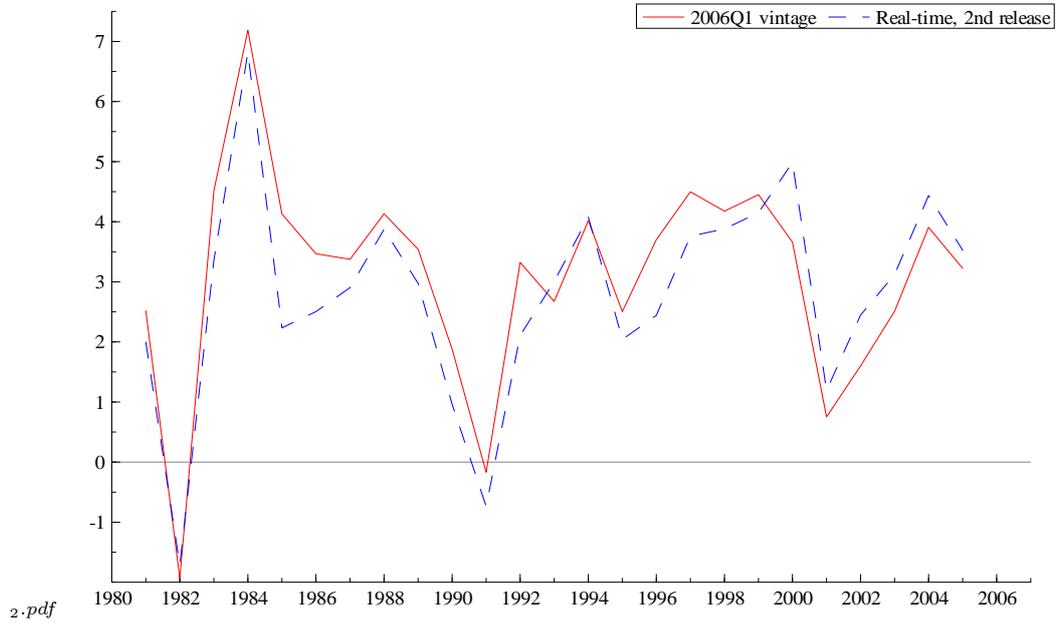


Figure 2: Annual real output growth 1981 to 2005 using the latest data vintage (2006:Q1) and the second release vintage ('real-time')

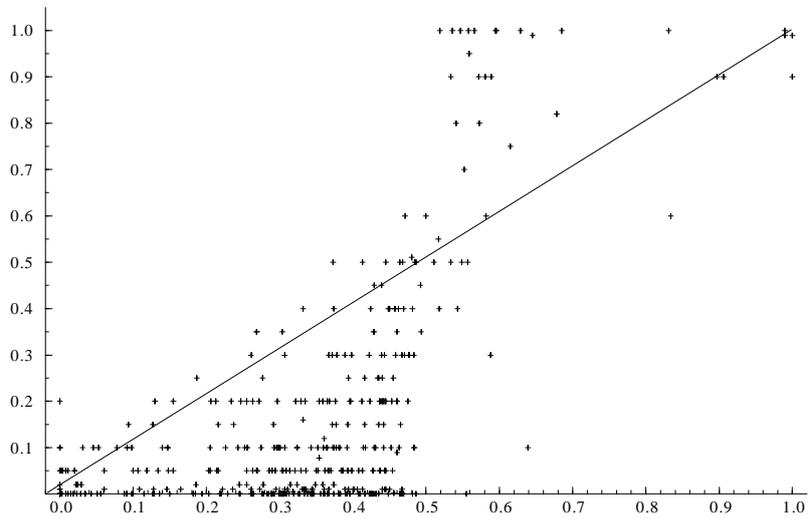


Figure 3: Scatter plots of current-quarter probability forecasts of decline and implied probabilities of decline from the point forecasts. Q4 surveys only, all individuals.