Subjective and Ex Post Forecast Uncertainty :
US Inflation and Output Growth

Michael P. Clements

No 995

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Subjective and Ex Post Forecast Uncertainty: US Inflation and Output Growth

Michael P. Clements*
Department of Economics,
University of Warwick,
UK. CV4 7AL.
m.p.clements@warwick.ac.uk.

September 5, 2012

Abstract

Survey respondents who make point predictions and histogram forecasts of macro-variables reveal both how uncertain they believe the future to be, *ex ante*, as well as their *ex post* performance. Macroeconomic forecasters tend to be overconfident at horizons of a year or more, but over-estimate the uncertainty surrounding their predictions at short horizons.

Journal of Economic Literature classification: C53.

Keywords: Subjective uncertainty, realized uncertainty, output growth forecasts, inflation forecasts.

*I am grateful to Malte Knüppel for helpful comments. Computations were performed using code written in the Gauss Programming Language.

1 Introduction

Uncertainty plays a key role in economic theories of agent behaviour. For example, uncertainty over future rates of inflation underpins many macroeconomic relationships, including Fisher equations for interest rates (see e.g., Lahiri, Teigland and Zaporowski (1988) and references therein), real wages (e.g., Ratti (1985)) and employment (e.g., Holland (1986)). Direct measures of agent uncertainty are rarely available, although the individual histograms of expected annual inflation and output growth reported by respondents to the US Survey of Professional Forecasters (SPF) over the last forty years have been recognized in the literature as a notable exception.[1] The histograms have been used, either individually or in aggregate, to derive direct measures of uncertainty that can serve as the gold standard for evaluating other ways of measuring or proxying forecast uncertainty. Chief amongst these are measures based on the dispersion of individuals' point forecasts ('disagreement' or 'consensus'): see Zarnowitz and Lambros (1987), with more recent contributions including Giordani and Söderlind (2003), Boero, Smith and Wallis (2008), D'Amico and Orphanides (2008) and Rich and Tracy (2010).

In this paper we analyze the properties of the histogram-based measures of uncertainty, and in particular the relationship between the histogram measure as an *ex ante* measure of forecast uncertainty, and *ex post* or realized uncertainty. Our interest in the term structure of forecast uncertainty (i.e., how forecast uncertainty changes with the horizon) is motivated in part by a recent paper by Patton and Timmermann (2011). Patton and Timmermann (2011) adopt a fixed-event framework to consider how realized forecast uncertainty varies over the forecast horizon. Their approach is silent about *ex ante* forecast uncertainty, and its relationship to *ex post* uncertainty. We use actual survey forecast errors to measure *ex post* uncertainty, as they do, but also use the SPF histograms to measure *ex ante* uncertainty.

The prevailing view is that agents' tend to be overconfident in their probability assessments, as indicated by the literature on behavioral economics and finance (see, e.g., the surveys by Rabin (1998) and Hirshleifer (2001)), which appears to be borne out by the limited empirical evidence that exists from surveys of macro-forecasters. Giordani and Söderlind (2003) construct confidence bands for the SPF respondents' inflation forecasts using the individual histogram standard deviations, and then compare the actual coverage to the nominal for three different levels. They consider the annual inflation forecasts from the first-quarter of the year surveys for the period 1969-2001, corresponding to an approximate one year-ahead horizon. They find

[1] Beginning in 1996, the Bank of England Survey of External Forecasters has provided similar information for the UK (see, for example, Boero, Smith and Wallis (2012)), and since 1999 the ECB Survey of Professional Forecasters (SPF) for the EURO area (see, e.g., Garcia (2003)).

overconfidence in that the actual coverage rates are markedly lower than the nominal. Giordani and Söderlind (2006) consider the US SPF real annual GDP (and GNP) forecasts 1982-2003, and again find overconfidence.[2] A recent study by Kenny, Kostka and Masera (2012) on the ECB's SPF also suggests overconfidence in the respondents' EURO area GDP growth and inflation forecasts at one and two-years ahead. Our findings suggest that US professional forecasters are not overconfident at the shorter horizons for our sample period, 1982-2010, for either inflation or output growth, and that on the contrary their subjective probability distributions clearly overstate the uncertainty characterizing the environments they are operating in.

Given the substantial heterogeneity in *ex ante* uncertainty found by Boero *et al.* (2012) in their study of UK forecasters (see also the panel approach of Lahiri and Liu (2006) on the US SPF), the study of the relationship between *ex ante* and *ex post* uncertainty needs to be made at the individual level. Nevertheless, consensus forecasts are often studied (see e.g., Ang, Bekaert and Wei (2007)) as are aggregate density functions (see, e.g., Diebold, Tay and Wallis (1999)), so we will also consider the relationship between *ex ante* and *ex post* uncertainty at the level of the consensus (or aggregate) forecasts.

The plan of the rest of the paper is as follows. Section 2 begins by setting out the relationship between EAU and EPU in an idealized setting, at the level of the data generating process, and then outlines our investigation of the relationship between EAU and EPU using survey data. Section 3 describes the data from the US Survey of Professional Forecasters (SPF). Section 4 investigates the relationship between EAU and EPU at the level of the consensus forecast, section 5 confirms earlier work suggesting a good deal of variation in EAU across individuals, which motivates the individual-level analysis reported in section 6. Finally, section 7 considers the extent to which our results are due to the recent financial crisis, given that the sample period runs from 1981 to 2010. Section 8 offers some concluding remarks.

2 Motivation

Simply for illustrative purposes, suppose the data generating process is:

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \quad \text{where } \varepsilon_t \sim D\left(0, \sigma^2_{\varepsilon,t}\right), \tag{1}$$

[2] They consider forecasts made in each of the four quarters of the year of the current year annual growth rate, i.e., forecasts from one year-ahead to one quarter-ahead (approxmately), and find 'strong consistency in coverage rates at all forecast horizons' so report coverage rates for all four quarters (i.e., horizons) jointly (see their table 2, p. 1035).

where $\sigma^2_{\varepsilon,t}$ follows an ARCH or GARCH process (say), then the true conditional forecast density of Y_t based on information through $t-h$ ($\Im_{t-h}$) is:

$$Y_{t|t-h} \sim N\left(\rho^h Y_{t-h}, \sigma^2_{y,t|t-h}\right) \tag{2}$$

where $E(Y_t \mid \Im_{t-h}) = \rho^h Y_{t-h}$, and $Var(Y_t \mid \Im_{t-h}) = \sigma^2_{y,t|t-h}$. The expected squared error of the optimal (MMSE) point prediction is of course:

$$E_{t-h}\left[\left(Y_t - Y_{t|t-h}\right)^2\right] = E_{t-h}\left[\left(\varepsilon_t + \rho\varepsilon_{t-1} + \ldots + \rho^{h-1}\varepsilon_{t-h+1}\right)^2\right] = \sum_{i=0}^{h-1}\rho^{2i}\sigma^2_{\varepsilon,t-i|t-h} = \sigma^2_{y,t|t-h}, \tag{3}$$

where $\sigma^2_{\varepsilon,t|t-h} = E\left(\sigma^2_{\varepsilon,t} \mid \Im_{t-h}\right)$, and we have assumed $E(\varepsilon_t\varepsilon_s) = 0$ for all $t \neq s$. Expression (3) is referred to as *ex post* forecast uncertainty (EPU), because in practice it is calculated by comparing the point prediction to the outcome. The variance in (2) is the *ex ante* forecast uncertainty (EAU), because it is an element of the individual's density *forecast* made prior to the realization of the outcome. At the level of the population, EPU and EAU coincide.

In practice individual respondents will not know the form of the data generating process (DGP) in (1), the DGP may be non-constant over time (see, e.g., Clements and Hendry (2006)), and individuals will approach the forecast problem with different beliefs about the appropriate model (or models, or forecasting methods) to use, and different expectations about the likely long-run values of the variables (see, for example, Lahiri and Sheng (2008) and Patton and Timmermann (2010)). Consequently, the equality between the EAU and the EPU at the population level is not necessarily a useful guide to empirical outcomes. Our interest is in the characteristics of EAU and EPU in practice, as revealed in survey data. Suppose individual i makes a conditional density forecast, and simultaneously issues a point prediction, both of Y_t, at each of a number of forecast origins, $t-1$, $t-2$, ... Moreover, the same respondent does so for each of a number of target periods (t). Let $y_{i,t|t-h}$ denote the point prediction, and $\sigma^2_{i,t|t-h}$ the variance of their density forecasts. Two considerations are of interest:

1. The relationship between their EAU, given by $\sigma^2_{i,t|t-h}$, and their EPU, given by comparing $y_{i,t|t-h}$ to the realizations. Is respondent i able to accurately foresee the uncertainty surrounding his/her point predictions? Does the relationship depend on the forecast horizon h?

2. Is it the case that individuals who think they are good forecasters (report lower than average EAUs) actually are (have lower than average EPUs), or alternatively, are such individuals simply overconfident?

These questions are one of the main the focuses of the paper.

3 The Survey of Professional Forecasters (SPF)

The SPF 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), Zarnowitz and Braun (1993) and Croushore (1993). The survey questions elicit information from the respondents on their point forecasts for a number of variables, including the level of real GDP and the GDP deflator, as well as their forecast distributions of the annual rate of change in these two variables, given in the form of histograms.[3]

We use the forecasts of annual output growth and inflation for the current year (i.e., the year of the survey) and the next year. The respondents report forecasts of the level of real output and the GDP deflator. The growth forecasts are constructed as follows. The current year forecast growth rates are the percentage changes between the forecasts of the annual level and latest estimate of the previous year's level available at the time of the survey.[4] The next years' growth rates are the percentage changes between the two annual levels forecasts. So a Q1 survey will provide a 4-step ahead forecasts of the current year's growth rates, and an 8-step ahead forecast of next period's growth rates, and a Q4 survey will provide 1-step ahead forecasts of the current year's growth rates, and a 4-step ahead forecast of next period's growth rates. Hence we have sequences of fixed-event forecasts with horizons of 8 down to 1 quarter for the annual growth rate in each year. We use the surveys from 1981:3[5] up to 2010:4, so have 1 to 8-step ahead forecasts of the 28 annual growth rates from 1983 to 2010. These annual growth forecasts are used to construct measures of *ex post* forecast uncertainty, where we use as actual values the second-release figures.[6]

The histograms refer to the annual change from the previous year to the year of the survey,

[3] The point forecasts and histograms have been widely analysed (see, *inter alia* Zarnowitz (1985), Keane and Runkle (1990), Davies and Lahiri (1999), for the point forecasts, and Diebold *et al.* (1999), Giordani and Söderlind (2003) and Clements (2006), for the probability distributions).

[4] The actual values are taken from the Real Time Data Set for Macroeconomists (RTDSM) maintained by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)). The RTDSMs contain the values of output that would have been available at the time the forecast was made, as subsequent revisions, base-year and other definitional changes that occurred after the reference date are omitted.

[5] Prior to this date, annual level forecasts for the current and next year were not recorded, and the histograms and point forecasts for output referred to nominal as opposed to real GDP(GNP). Further, the inflation histograms for some of the earlier periods did not always refer to the current year (i.e., survey quarter year) and the following year.

[6] These are again taken from the RTDSM. A number of authors have recently used the 'second-release' data vintage series as 'actual values', by which we mean the vintage available two quarters later (see e.g., Romer and Romer (2000)).

as well as of the survey year to the following year, allowing the construction of sequences of fixed-event histogram forecasts that match the annual growth rate forecasts in timing and number of horizons. The difficulties of calculating second moments from the histograms as measures of *ex ante* forecast uncertainty have been documented in the literature. The basic problem arises because the histogram is a discretized version of the individual's probability distribution (which is assumed to be continuous) and that the moments of interest are of the continuous distribution. Following Zarnowitz and Lambros (1987), it is generally assumed that the open-ended first and last intervals are closed with width equal to that of the inner bins, and then that either the probability mass is uniform within each bin or is located at the mid-points of the bins. We adopt the mid-points assumption, and use Sheppard's correction (see e.g., Heitjan (1989)), and let σ_{ith}^{m} denote the resulting estimate of the standard deviation of the histogram of individual i made at time t with a horizon of h quarters. Alternatively, one can assume a parametric form for the distribution underlying the histogram. Giordani and Söderlind (2003) fit normal distributions to the histograms as a way of estimating the variance, and Engelberg, Manski and Williams (2009) fit generalized beta distributions. We follow Boero *et al.* (2012) and fit triangular distributions when probabilities are assigned to only one or two histogram intervals, and fit normal distributions when three or more intervals have non-zero probabilities. We let σ_{ith}^{p} denote an estimate obtained in this fashion.

4 A comparison of subjective and *ex post* forecast uncertainty for the survey consensus

The consensus forecasts are often analyzed as a convenient summary of the information in the survey. They also permit use of all the quarterly SPF surveys, whereas the number of responses for most individuals is far fewer. The EPU measure is based on the consensus point forecasts, calculated as the cross-section average of the respondents' forecasts. In the context of the SPF forecasts 1968-1990, Zarnowitz and Braun (1993) show that the consensus forecasts 'have considerably smaller errors than the average individual respondent' (p.36), while Manski (2011) provides the analytic support for such findings. We also report the average of the individuals' squared forecast errors for comparison purposes.

Our EAU measure is the average of the individuals' histogram standard deviations. Standard results show that the variance of the aggregate histogram equals the average variance plus the (variance-measure) of the dispersion of the mean forecasts, so that using the standard deviation of the aggregate histogram would inflate EAU relative to our measure (see, e.g., Giordani and Söderlind (2003) and Wallis (2005)). We do not report the standard deviation of the aggregate histogram because our results show that the average of the individual standard deviations is itself large relative to EPU at the within-year horizons, so that using the standard deviation of the aggregate histogram would only exacerbate the mismatch between the EAU and EPU measures.

Formally, the estimate of the consensus EPU $\hat{\sigma}_{h,ep}$ is the sample standard deviation of the consensus forecast errors at horizon h ($h = 1, \ldots, 8$):

$$\hat{\sigma}_{h,ep} = \sqrt{T^{-1} \sum \left(e_{t|t-h} - e_h\right)^2}$$

where $e_{t|t-h} = y_t - y_{t|t-h}$, $e_h = T^{-1} \sum e_{t|t-h}$, $y_{t|t-h} = N_t^{-1} \sum_{i=1}^{N_t} y_{i,t|t-h}$, and N_t is the number of respondents who forecast period t (typically this will vary a little with the horizon, h, but we suppress this for simplicity). $T = 28$. We also report the consensus RMSE given by:

$$RMSE_{h,ep} = \sqrt{T^{-1} \sum e_{t|t-h}^2}.$$

The average EPU across individuals is given by:

$$\overline{\sigma}_{h,ep} = N^{-1} \sum_{i=1}^{N} \hat{\sigma}_{i,h,ep} = N^{-1} \sum_{i=1}^{N} \sqrt{T_i^{-1} \sum \left(e_{i,t|t-h} - e_{i,h}\right)^2},$$

where $e_{i,t|t-h} = y_t - y_{i,t|t-h}$, $e_{i,h} = T_i^{-1} \sum e_{i,t|t-h}$, T_i is the number of forecasts by individual i, and similarly for the average RMSE, which we denote $\overline{RMSE_{h,ep}}$.

The subjective measures are given by $\sigma_h^k = T^{-1} \sum_t \left(N_t^{-1} \sum_i \sigma_{i,t|t-h}^k\right) \equiv T^{-1} \sum_t \sigma_{t|t-h}^k$, for $k \in \{m, p\}$, so that for a given horizon h ($h = 1, \ldots, 8$), we average over all individuals to give the cross-section average ($\sigma_{t|t-h}^m$, $\sigma_{t|t-h}^p$), and then average over time (t).

Table 1 reports the term structure of EPU (in standard deviation form) and EAU based on the consensus forecasts. For EAU we report the standard histogram measure, σ_h^m as well as the estimate obtained using the normal distribution (and triangular distributions, as appropriate), σ_h^p. These essentially tell the same story, with σ_h^p tending to be a little smaller, as expected, and we focus on σ_h^p. For EPU we report the standard deviation of the consensus forecasts $\widehat{\sigma}_{h,ep}$ as well as the RMSFE, which show similar profiles over h. For all four measures we report scaled versions, where in each case we divide the whole column by the $h = 8$ value. These are given in the columns headed by 's.'. For example, for σ_h^p, the adjacent column 's.' is given by σ_h^p / σ_8^p.

Comparing σ_h^p to $\widehat{\sigma}_{h,ep}$ shows that for both inflation and output growth EAU clearly overstates EPU as the horizon shortens. The consensus forecast EAUs are far too pessimistic at short horizons. For output growth, EAU understates the actual uncertainty surrounding the consensus forecasts at the longer horizons (in excess of one year). Note that the use of the standard deviation of the aggregate histogram would further exacerbate the difference at the shorter horizons. For output growth, there are large reductions in realized uncertainty between the $h = 5$ and $h = 4$ horizons, and between $h = 4$ and $h = 3$. These correspond to moving from a Q4 survey of next year's output growth to a Q1 survey of this year's output growth, and moving from a Q1 to Q2 survey forecast of current year output growth. Evidently there are marked improvements in the accuracy of consensus forecasts when there is some information on the first quarter of the year being forecast, and on the second quarter of the year,[7] although the 'carry-over effect' would imply a similar pattern (see below). On the basis of these estimates, an approximate 95% interval for the annual rate of output growth made in the fourth quarter of that year would be roughly the central projection $\pm\frac{1}{2}$% point, whereas the *perceived* interval would be ± 1% points.

For inflation, EAU is a closer match to EPU at long horizons, but otherwise the profile over h is the same: EAU remains high relative to EPU as h shortens. EPU again declines markedly going from $h = 5$ to $h = 4$. At $h = 1$, EAU is roughly two and a half times as large as EPU. For both variables the term structure profile of EAU displays less responsiveness to current-year information.

The EPU RMSFE figures are similar to those for the standard deviation for output growth, because the consensus forecasts are largely unbiased. For inflation there are marked biases at the longer horizons. At $h = 7$, for example, the consensus forecasts over-estimate inflation by nearly half a percentage point on average. Nevertheless, whether we use the EPU standard deviation, or incorporate bias using the RMSE, the same picture emerges - EAU uncertainty remains high relative to the EPU measure.

We also report the averages of the individuals' EPU forecast standard deviations and RMSEs. As expected, these are a little less accurate than the corresponding consensus measures, but display the same profile over h.

Table 1 also records a 'theoretical' forecast uncertainty measure which provides a profile for uncertainty as h changes, based on some simple but reasonably plausible assumptions about the predictably of quarterly growth rates. Note that the year-on-year annual growth rate can be closely approximated by a weighted average of the relevant quarter-on-quarter growth rates.

[7] Note that the timing of the SPF (middle of the second month of the quarter) is such that responses to Q1 survey forecasts will be informed by the first estimate of non-farm payroll for the first month of the quarter (i.e., for January).

If we let q_t be the first-difference of the log-level of the variable (output, or the price deflator) in the fourth-quarter of year t, then the year-on year growth rate, y_t, is approximately given by $y_t = \frac{1}{4}\sum_{j=0}^{3}\left(\sum_{s=0}^{3} q_{t-j-s}\right)$, because $\sum_{s=0}^{3} q_{t-j-s}$ is the annual growth rate between $t-j$ and $t-j-4$, etc. Rearranging, we can write this as $y_t = \sum_{j=0}^{6} w_j q_{t-j}$ where $w_j = \frac{j+1}{4}$ for $0 \le j \le 3$, and $w_j = \frac{7-j}{4}$ for $4 \le j \le 6$. If we assume that the q_t are iid, with $Var\left(q_t\right) = \sigma_q^2$, then forecast uncertainty at horizon h is given by:

$$\sigma_h^2 \equiv Var\left(y_t \mid q_{t-h}, q_{t-h-1}, \ldots\right) = \sigma_q^2 \sum_{j=0}^{h-1} w_j^2 \tag{4}$$

where $h = 1$ to 8. Note that $h = 1$ corresponds to a forecast made in the fourth quarter of the year of the annual (year-on-year) rate of growth for that year, whereas $h = 8$ corresponds to a forecast made in the first-quarter of the previous year. When $h = 1$, for example, $\sigma_1^2 = w_0^2\sigma_q^2$, as the only unknown component is q_t. The key point is that weights w_j on the quarterly growth rates are 'tent-shaped', with w_3 a maximum (referring to the first quarter of the year being forecast), so that uncertainty does not decline linearly as the forecast horizon shortens. In particular, there ought to be a large reduction in forecast uncertainty between the second and first quarter once the first quarter growth becomes known. This is sometimes referred to in the literature as the 'carry-over effect' (see Tödter (2010) for an exposition and an empirical analysis of forecasting German real GDP). Patton and Timmermann (2011) allow for a more elaborate model which allows that the q_t have a persistent, predictable component, so that the q_t are no longer *iid*. We continue with the simpler formulation which results in (4). In table 1 the column headed 'Theor.' is σ_h/σ_8, so that the measure does not depend on σ_q^2 (although this could be estimated from the data), and is directly comparable to the 's.' columns. Note that forecast uncertainty in the second quarter ($h = 3$) is 56% of the initial level (of the standard deviation), compared to 83% in the first quarter ($h = 4$) when the quarterly value for the first quarter is not known. For both variables the decline in $\widehat{\sigma}_{h,ep}$ from $h = 8$ to $h = 1$ is broadly similar to that predicted by (4), from 1 to 0.18 and 0.25 for output growth and inflation respectively, compared to 0.15 for the theoretical measure. The profile over h of EPU for both variables is lower than the theoretical value at medium horizons, indicating that such forecasts exploit additional information over and above that assumed in (4), so that quarterly growth rates are predictable beyond the unconditional mean. However, the principle value of (4) is to show that EA uncertainty is 'too high' at the within-year horizons, as it remains higher than indicated by (4) for all $h < 4$.

Our findings are in tune with those of Diebold *et al.* (1999) and Clements (2006), although those studies do not consider the term structure. They calculate probability integral transforms for the SPF aggregate histograms made in the first quarters of each year, and show that the histograms appear to overstate the uncertainty surrounding inflation in more recent times. Their forecasts correspond to our $h = 4$ horizon, at which we find the EAU is roughly twice the EPU for inflation. Giordani and Söderlind (2003) argue that the use of the aggregate histogram in those studies will lead to uncertainty being overstated compared to a representative individual's uncertainty. As table 1 uses the average individual standard deviation, rather than the standard deviation of the aggregate histogram, it is immune to this criticism. Giordani and Söderlind (2003) construct confidence bands for the SPF respondents' point forecasts using the individual histogram standard deviations, and then compare the actual coverage to the nominal. Their results indicate that forecasters underestimate uncertainty, because the actual coverage rates are less than the nominal (for the period 1968 - 2000), but results are only reported for the first quarter surveys of the current year (corresponding to $h = 4$) and so again provide no information on the term structure.

We conclude that, based on either the consensus forecasts or the average forecast performance of the respondents, the EAU surrounding future output growth and inflation remains higher than is warranted as the horizon shortens. In the following we investigate whether some respondents' perceptions are more in tune with the actual uncertainty they face.

5 Disagreement about subjective uncertainty and differences in forecast accuracy

A number of authors have studied disagreement about perceived forecast uncertainty (see D'Amico and Orphanides (2008) and Boero *et al.* (2012)), and others the extent to which forecasters differ in terms of their realized forecast performance (see D'Agostino, McQuinn and Whelan (2012)). In this section we ask whether the differences in perceived uncertainty across individuals matches actual differences in forecast accuracy across individuals.

We measure the degree of disagreement about perceived uncertainty as the standard deviations across individuals' horizon-specific EAU measures, where the individual measures are time-averages of uncertainty estimates obtained from the histogram forecasts, namely, $\sigma^p_{ih} = T^{-1} \sum_{t=1}^{T} \sigma^p_{ith}$ (where the value of 'T' typically depends on i and h). We calculate σ^p_{ih} for each i for which we have at least 5 recorded histograms (for that h). Table 2 records the standard deviations of the σ^p_{ih} over i as a measure of disagreement about EAU. In addition, we report the cross-sectional standard deviations of the *ex post* standard deviations, so that the dispersion of actual skill levels is compared to the dispersion of perceived skill levels.

We find that the cross-sectional dispersion of perceived uncertainty is markedly lower than the dispersion of actual forecast performance at the two-year forecast horizon. As the forecast horizon shortens, the dispersion of actual performance declines roughly monotonically to around a third of the two-year ahead level at $h = 1$. But the dispersion of perceived uncertainty remains high, and at $h = 2$ is nearly as high as at $h = 8$, for both output growth and inflation. We also record the cross-sectional standard deviations of respondents' RMSFEs (as opposed to the dispersion of their forecast-error standard deviations). These are broadly similar to the results for forecast standard deviations, so that the results are not qualitatively affected by individual-level biases.

It should be noted that the levels of the dispersion of both perceived uncertainty and *ex post* forecast uncertainty are likely to be inflated by the unbalanced nature of our panel of forecasters. Our sample of SPF data covers nearly thirty years, so that respondents will have entered and exited the survey over this period (in addition to occasional non-responses from otherwise active participants).[8] Hence they will have reported their forecasts during potentially quite different economic conditions. However, whilst this will affect the level of the dispersions, it should not affect the relationship between the dispersion of perceived and *ex post* forecast uncertainty, or their shapes as the forecast horizon shortens. This is because a respondent active during a relatively tranquil (volatile) period will report lower (higher) than average perceived uncertainty, but also lower (higher) than average *ex post* forecast uncertainty.

We have shown that the term structure of the realized accuracy of the consensus forecasts is such that the shorter horizon forecasts are markedly more accurate than indicated by the (average) *ex ante* uncertainty. Furthermore, the dispersion of perceived uncertainty across individuals fails to reflect the actual differences in forecast ability between individuals at the longer (18 month to two-year) horizons, and is largely insensitive to the horizon (apart from at the shortest, one-quarter horizon).

6 Assessing individuals' assessments of forecast uncertainty

In this section we ask whether individual macro-forecasters are able to accurately assess (*ex ante*) the uncertainty they face (*ex post*). Is it the case that a forecaster who perceives a low level of uncertainty (a relatively confident forecaster) has correspondingly more accurate

[8] For surveys spanning a shorter historical period, such as the ECB SPF (1999 to the present), it is possible to obtain a balanced panel by dropping some respondents and 'filling in' missing observations - see Genre, Kenny, Meyler and Timmermann (2010) and Kenny *et al.* (2012). Capistrán and Timmermann (2009) consider entry and exit in the context of forecast combination.

forecasts?

For each individual with eight or more forecasts at each of the eight forecast horizons we estimate measures of *ex ante* and *ex post* uncertainty. We adapt the estimates described in sections 4 and 5 to counter potential distortions from the unbalanced nature of our panel of forecasts: some respondents might have faced predominantly easier (harder) conditions than others, so that a forecaster with a lower EPU is not necessarily a 'better' forecaster. To calculate EPU, the actual squared forecast errors are weighted by the cross-sectional average for that t relative to the average over all t. Thus, if the average (over forecasters) squared forecast error at period t was large relative to forecasts made at other times, the squared errors of all who forecast period t will be scaled down.

Formally, the weighted RMSE is given by:

$$RMSE^{*}_{i,h,ep} = \sqrt{T_i^{-1} \sum_{t}^{T_i} e^{*2}_{i,t|t-h}} \tag{5}$$

where

$$e^{*2}_{i,t|t-h} = e^{2}_{i,t|t-h} \times \frac{median_t(median_i(|e_{i,t|t-h}|))}{median_i(|e_{i,t|t-h}|)}$$

where we take the absolute value of the forecast errors rather than the squares, and the median rather than the mean to lessen dependence on outliers. $median_i\left(\right)$ is the cross-sectional median, and $median_t\left(\right)$ is the median over t. We also calculate a 'standard deviation' measure as

$$\sqrt{T_i^{-1} \sum_{t}^{T_i} e^{*2}_{i,t|t-h} - \left(T_i^{-1} \sum_{t}^{T_i} e_{i,t|t-h}\right)^2} \tag{6}$$

For the *ex ante* measure, we proceed similarly

$$\sigma^{*}_{i,h} = T_i^{-1} \sum_{t}^{T} \sigma^{*}_{i,t|t-h} \tag{7}$$

where

$$\sigma^{*}_{i,t|t-h} = \sigma_{i,t|t-h} \times \frac{\frac{1}{T}\sum_{t}^{T} median_i(\sigma_{i,t|t-h})}{median_i(\sigma_{i,t|t-h})}$$

so that the time-average is the mean. We also report results without any weighting to see whether the results are qualitatively affected by weighting for the ease of forecasting.

Figures 1 and 2 present scatter plots of $RMSE^{*}_{i,h,ep}$ (y-axis) against $\sigma^{*}_{i,h}$ (x-axis) for output growth and inflation, respectively. Points that lie on the 45-degree line indicate individuals whose subjective assessments match outcomes. Points above the 45% line denote overconfidence, and those below underconfidence. There is little evidence of a positive relationship between EPU and EAU at any horizon for output growth, and at within-year horizons clear evidence of underconfidence. The story for inflation is essentially the same, although at the longer horizons it appears that there might be a negative relationship - those who believe themselves worse are in fact more accurate!

What happens if we do not attempt to adjust for the ease/difficulty of forecasting? The story for output growth is largely unchanged at the within-year horizons, although at the longer horizons there is more evidence of overconfidence (see figure 3), but no more or less indication that EAU and EPU are positively related. For inflation there is little discernible effect (see figure 4).

Table 3 reports Spearman rank order correlation coefficients between the individuals' EAU and EPU estimates as a more formal indication of whether there is a monotonic relationship between *ex ante* and *ex post* forecast uncertainty. There is little evidence of a positive monotonic relationship for either variable at any horizon, irrespective of whether an adjustment is made for the unbalanced nature of the panel. For inflation there is some evidence of a negative relationship at longer horizons, especially when adjustments are made, bearing out the visual impression from the scatterplots.

In summary, the following findings hold irrespective of whether adjustments are made for the unbalanced nature of the panel: a) there is little evidence that more (less) confident forecasters are more (less) able forecasters, and b) individuals are underconfident at shorter (within-year) horizons.

Whether individuals are over or underconfident is often addressed by comparing actual coverage rates of prediction intervals to their nominal levels (see, e.g., Giordani and Söderlind (2003, 2006)), so we report interval coverage rates to enable comparison of our findings to those in the literature, and as a complement to the results on EPUs and EAUs. Prediction intervals may have smaller coverage rates than intended because the interval is located in the wrong place rather than the 'scale' being wrong (Giordani and Söderlind (2006) refer to this as optimism or pessimism versus 'doubt'). Our base case prediction intervals are calculated by fitting normal approximations to the individual histograms. We then consider two alternatives to correct for location errors. In the first, we re-centre the intervals on the histogram means bias-corrected for each individual (for a given h). If a respondent tends to be pessimistic about growth prospects, in the sense that the means of his histograms are systematically lower than the outcomes, the intervals are moved to the right. Secondly, individuals' histogram means and point predictions are not always consistent one with another (see Engelberg *et al.* (2009)), and Clements (2009,

2010) shows that the point predictions tend to be more accurate (under squared-error loss). For this reason, we also report results for intervals located on the point predictions, and intervals centred on the bias-corrected point predictions.

Table 4 reports actual coverage rates for two nominal levels (50% and 90%). The coverage rates are calculated across all individuals and time periods for a given h (the number of intervals in each case is recorded in the columns headed '#'). Consider firstly the results for output growth. When the intervals are centred on the mean, we find that actual coverage rates are too low for both the 50% and 90%-intervals at the longer horizons, but the intervals are closer to being correctly-sized for shorter horizons. When the intervals are centred on the point forecasts, they are clearly over-sized at the within-year horizons, indicating underconfidence. Bias correction of the histogram means or point predictions does not qualitatively affect the results. For inflation a similar story emerges - there is clear evidence of underconfidence at within-year horizons when the intervals are centred on the point predictions. At $h = 1$, for example, the nominal 50% inflation interval has an actual coverage rate of 90%; that of a 90% inflation interval a coverage rate of 96%.

Finally, we calculate formal tests of whether the respondents' subjective assessments are in tune with the *ex ante* outcomes. We consider all those who responded to ten or more surveys of a given quarter of the year. For each horizon h and respondent i, we directly compare the *ex ante* and *ex post* uncertainty assessments by calculating $w_{i,t|t-h} = e_{i,t|t-h}/\sigma_{i,t|t-h}$, and then test whether $E\left(w_{i,t|t-h}^2\right) = 1$.[9] We regress $w_{i,t|t-h}^2$ on a constant, and test the hypothesis that the constant is one. We consider one-sided alternatives - if we reject the null in favour of the constant being less than one, then we conclude that respondent i at horizon h is prone to underconfidence (their subjective assessments of uncertainty exceed the actual uncertainty they face). Similarly, rejecting in favour of the constant exceeding one indicates overconfidence. Rather than reporting the results separately for each h, we take all the within-year forecasts together, and all the next-year forecasts together. The next-year forecasts are multi-step ahead forecasts in the sense that a forecast is made before the outcome corresponding to the previous forecast is known.[10] This gives rise to the well-known point forecast problem of forecast errors

[9]Formally, the w's correspond to inverse-normal probability integral transforms (IN-PITs) (see Berkowitz (2001) and Knüppel (2011) for a review) when the forecast density is gaussian, and $e_{i,t|t-h}$ is the outcome minus the mean (rather than the point prediction). However, our goal is less ambitious that a PIT evaluation, as our interest is in forecast uncertainty, rather than whether the forecast densities are correctly-calibrated. We fit normal distributions to the histograms as a way of estimating the variances of the histograms, supposing that the estimator of the variance based on fitting a normal distribution to a histogram is a reasonably robust estimator even if the distributional assumption is incorrect. A PIT evaluation assesses whether the distributional assumption is correct throughout its range.

[10]This is not true of the within-year forecasts except to the extent that we use 'final' actuals - those available

being correlated, so that the w's are correlated, and we use autocorrelation-consistent standard errors. We also report tests which adjust for potential bias in the point forecasts. In that case, we replace $w_{i,t|t-h} = e_{i,t|t-h}/\sigma_{i,t|t-h}$ by $w_{i,t|t-h} = (e_{i,t|t-h} - e_{i,h})/\sigma_{i,t|t-h}$, where $e_{i,h}$ is the sample mean of the forecast errors. If the forecasts are biased, this adjustment means we are comparing the *ex post* forecast standard deviation - rather than the RMSFE - with the *ex ante* forecast standard deviation. Note that this correction will not necessarily bring a better match between EA and EP uncertainty - if EA over-estimates EP, removing the bias from EP would exacerbate the mis-match.

Table 5 shows that we tend to reject the null in favour of underconfidence at the within-year horizons for both variables, but in favour of overconfidence at the longer horizons, as might have been anticipated from the results for interval coverage rates (see table 4). For inflation, the rejection rate for the within-year forecasts is around two thirds (when bias adjusted), and around one third for output growth, while the next-year rejection rates in favour of overconfidence are higher for output growth than inflation. Given that these tests might be expected to have low power given the small sample sizes (average number of observations per regression is 14), the fact that we reject the null as often as indicated in the table casts doubt on whether individuals are able to accurately assess the uncertainty they face.

7 The influence of the financial crisis

The recent recession was one of unprecedented severity, and its inclusion might be expected to have inflated the realized forecast errors underpinning the EPU calculations. If the recent recession is viewed as a special event, of interest is the extent to which it might have distorted our findings. Hence we exclude the forecasts of 2008 and later years, and repeat some of the calculations using only the surveys forecasts up to 2006:4. To save space, we report table 6, which reproduces the table 1 the results for the consensus forecasts. The individual-level results were largely unaffected, as the majority of respondents were not active during this period.

Table 6 shows that excluding the 'recession surveys' has little affect on the consensus EAU and EPU figures for inflation. As expected, there is a marked decrease in output growth realized uncertainty (compare $\hat{\sigma}_{h,ep}$ or RMS between tables 1 and 6) at the longer horizons, but the key finding that EAU remains high relative to EPU as h shortens remains. Indeed, the EAU for output growth is hardly affected by the recession surveys at any h, indicating that the dramatic

two quarters later - so that the first-quarter survey forecasts are overlapping. For example, the 2005:Q1 forecast of the current year is made before the 'final' value of the 2004 annual growth rate / inflation rate is released.

changes in realized uncertainty that characterize the end of the sample are not matched by changes in perceived uncertainty.

8 Conclusions

The conventional view that agents' tend to be overconfident in their probability assessments is not true of our sample of US professional macro-forecasters at horizons of up to one year ahead. Our panel of forecasters is in aggregate underconfident at the within-year horizons, as are many of the individual respondents, in the sense that the outlook for inflation and output growth is less uncertain that they perceive it to be. At the longer horizons, in excess of one year, there is tendency to overconfidence. By comparing subjective and objective uncertainty over horizons ranging from 2-years' ahead down to one-quarter ahead, we map out how actual uncertainty changes with the horizon (as in Patton and Timmermann (2011)) and compare its evolution with agents' perceptions of the uncertainty they face. The key difference between *ex ante* and *ex post* uncertainty is the tendency of the subjective measure to remain at a high level compared to the realized measure as the forecast horizon shortens. This is true of the consensus forecast errors and the average of the individual respondents forecast standard deviations, as well as of individual respondents. One interpretation of this phenomenon is that forecasters are unaware of the so-called 'carry-over' effect, whereby knowledge of the quarterly growth rates (or equivalently, levels of output or price levels) revealed as the forecast horizon shrinks engenders sharp reductions in uncertainty about the annual year-on-year growth rate, especially as current year quarterly values become available to the forecaster. We provide some illustrative calculations of this for the simplest case of uncorrelated quarterly growth rates, but the same phenomenon holds more generally.

There is little evidence to support the hypothesis that respondents who believe the future is less uncertain are more capable forecasters *ex post* than those who express the belief that the future is relatively more uncertain. There is a good deal of heterogeneity across forecasters both in terms of *ex post* forecast accuracy and in terms of *ex ante* subjective assessments, but there does not appear to be a systematic relationship between the two.

Finally, our overall findings are not unduly influenced by the recent financial crisis. If we omit the surveys in 2007 to 2010 the results are qualitatively unchanged.

References

Ang, A., Bekaert, G., and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better?. *Journal of Monetary Economics*, **54**, 1163–1212.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, **19**, 465–474.

Boero, G., Smith, J., and Wallis, K. F. (2008). Uncertainty and disagreement in economic prediction: the Bank of England Survey of Economic Forecasters. *Economic Journal*, **118**, 1107–1127.

Boero, G., Smith, J., and Wallis, K. F. (2012). The measurement and characteristics of professional forecasters' uncertainty. Discussion paper.

Capistrán, C., and Timmermann, A. (2009). Forecast combination with entry and exit of experts. *Journal of Business & Economic Statistics*, **27**, 428–440.

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.

Clements, M. P. (2009). Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Chapter 8*, pp. 206–226. Oxford: Oxford University Press.

Clements, M. P. (2010). Explanations of the Inconsistencies in Survey Respondents Forecasts. *European Economic Review*, **54**, 536–549.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 605–657: Elsevier, Horth-Holland.

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**, 111–130.

D'Agostino, A., McQuinn, K., and Whelan, K. (2012). Are some forecasters really better than others?. *Journal of Money, Credit and Banking*, **44**, 715–732.

D'Amico, S., and Orphanides, A. (2008). Uncertainty and disagreement in economic forecasting. Finance and economics discussion series 2008-56, Board of Governors of the Federal

Reserve System (U.S.).

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.

Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999). 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.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, **27**, 30–41.

Garcia, J. A. (2003). An introduction to the ECB's Survey of Professional Forecasters. mimeo, Occasional Paper Series No. 8, European Central Bank.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2010). Combining the forecasts in the ECB SPF: Can anything beat the simple average?. Working paper series no. 1277, European Central Bank.

Giordani, P., and Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, **74**, 1037–1060.

Giordani, P., and Söderlind, P. (2006). Is there evidence of pessimism and doubt in subjective distributions? implications for the equity premium puzzle. *Journal of Economic Dynamics & Control*, **30**, 1027–1043.

Heitjan, D. F. (1989). Inference from grouped continuous data: a review. *Statistical Science*, **4**, 164–183.

Hirshleifer, F. (2001). Investor psychology and asset pricing. *Journal of Finance*, **56**, 1533–1597.

Holland, A. S. (1986). Wage indexation and the effect of inflation uncertainty on employment: An empirical analysis. *American Economic Review*, **76**, 235–243.

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.

Kenny, G., Kostka, T., and Masera, F. (2012). How informative are the subjective density forecasts of macroeconomists?. Working paper series no. 1446, European Central Bank.

Knüppel, M. (2011). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. Discussion paper series 1: Economic studies, 32/2011, Deutsche Bundesbank, Research Centre.

Lahiri, K., and Liu, F. (2006). Modelling multi-period inflation uncertainty using a panel of density forecasts. *Journal of Applied Econometrics*, **21**, 1199–1219.

Lahiri, K., and Sheng, X. (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics*, **144**, 325–340.

Lahiri, K., Teigland, C., and Zaporowski, M. (1988). Interest rates and the subjective probability distribution of inflation forecasts. *Journal of Money, Credit and Banking*, **20**, 233–248.

Manski, C. F. (2011). Interpreting and combining heterogeneous survey forecasts. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting, Chapter 16*, pp. 457–472. Oxford: Oxford University Press.

Patton, A. J., and Timmermann, A. (2010). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, **57**, 803–820.

Patton, A. J., and Timmermann, A. (2011). Predictability of output growth and inflation: A multi-horizon survey approach. *Journal of Business and Economic Statistics*, **29**, 397–410.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, **36**, 11–46.

Ratti, R. A. (1985). The effects of inflation surprises and uncertainty on real wages. *Review of Economics and Statistics*, **67**, 309–314.

Rich, R., and 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.

Romer, C. D., and Romer, D. H. (2000). Federal Reserve information and the behaviour of interest rates. *American Economic Review*, **90**, 429–457.

Tödter, K.-H. (2010). How useful is the carry-over effect for short-term economic forecasting?. Discussion paper series 1: Economic studies, 21/2010, Deutsche Bundesbank, Research Centre.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, **67**, 983 – 994.

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 Braun, P. (1993). Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In Stock, J., and Watson, M. (eds.), *Business Cycles, Indicators, and Forecasting*, pp. 11–84: Chicago: University of Chicago Press and NBER.

Zarnowitz, V., and Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, **95**, 591–621.

Table 1: Consensus forecasts: Subjective and *Ex Post* Inflation and Output Growth Forecast Uncertainty.

h	Theor.	Annual Output Growth										Annual Inflation									
		Subjective (EA)				*Ex Post*		*Ex Post*		Average		Subjective (EA)				*Ex Post*		*Ex Post*		Average	
		σ_h^m	s.	σ_h^p	s.	$\widehat{\sigma}_{h,ep}$	s.	RMS	s.	EPU	RMS	σ_h^m	s.	σ_h^p	s.	$\widehat{\sigma}_{h,ep}$	s.	RMS	s.	EPU	RMS
8	1.00	1.10	1.00	1.01	1.00	1.60	1.00	1.61	1.00	1.66	1.79	0.89	1.00	0.85	1.00	0.75	1.00	0.84	1.00	0.86	0.97
7	1.00	1.06	0.97	0.99	0.98	1.55	0.96	1.55	0.96	1.53	1.62	0.91	1.02	0.87	1.03	0.65	0.87	0.81	0.96	0.86	1.00
6	0.99	1.03	0.94	0.95	0.94	1.39	0.87	1.39	0.87	1.55	1.63	0.86	0.96	0.84	0.99	0.58	0.78	0.72	0.86	0.76	0.88
5	0.94	0.96	0.87	0.88	0.87	1.14	0.71	1.15	0.72	1.23	1.29	0.80	0.89	0.78	0.91	0.51	0.68	0.56	0.67	0.67	0.78
4	0.83	0.91	0.83	0.83	0.82	0.80	0.50	0.86	0.54	0.80	0.90	0.79	0.88	0.73	0.86	0.36	0.48	0.36	0.43	0.53	0.57
3	0.56	0.82	0.75	0.76	0.75	0.53	0.33	0.53	0.33	0.57	0.56	0.74	0.83	0.70	0.83	0.27	0.36	0.30	0.36	0.37	0.40
2	0.34	0.70	0.64	0.63	0.63	0.39	0.24	0.39	0.24	0.41	0.41	0.66	0.74	0.61	0.72	0.22	0.29	0.23	0.27	0.31	0.31
1	0.15	0.55	0.50	0.46	0.46	0.28	0.18	0.28	0.18	0.30	0.29	0.53	0.60	0.48	0.57	0.19	0.25	0.19	0.23	0.22	0.22

Notes. The estimates are based on the surveys from 1982:1 to 2010:4, of annual output growth and inflation in 1983 to 2010 (28 years), averaged across respondents and surveys for a given horizon. σ_h^m and σ_h^p indicate that the standard deviations of the histograms are calculated using the standard formula and a normal approximation. The EPU standard deviation $\widehat{\sigma}_{h,ep}$ and the RMSFE ('RMS') use the second-release real-time data series to calculate forecast errors using the consensus annual point forecasts. The average EPU and RMS, denoted as $\overline{\sigma}_{h,ep}$ and $\overline{RMSE_{h,ep}}$ respectively, in the main text, are the cross-section averages of the individual EPU standard deviations and RMSEs. The columns headed 's.' scale the adjacent left columns by the $h = 8$ value to aid comparison with the theoretical uncertainty values ('Theor.'), which are unity for $h = 8$.
The calculations omit point forecasts made in the 85:1, 86:1 and 90:1 surveys, and histograms from 85:1 and 86:1, as it is not clear that these forecasts are comparable: see http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf.

Table 2: Disagreement about *ex ante* uncertainty and dispersion of realized uncertainty

	Annual Output Growth				Annual Inflation			
Horizon	No. of Individuals	Standard deviation *Ex ante*	Standard deviation *Ex post*	RMSE	No. of Individuals	Standard deviation *Ex ante*	Standard deviation *Ex post*	RMSE
8	53	0.30	0.50	0.57	54	0.29	0.58	0.53
7	66	0.37	0.44	0.49	65	0.32	0.48	0.53
6	64	0.39	0.49	0.46	64	0.35	0.33	0.36
5	47	0.34	0.39	0.37	52	0.34	0.28	0.32
4	53	0.24	0.29	0.33	54	0.24	0.25	0.26
3	66	0.27	0.20	0.20	65	0.26	0.27	0.25
2	64	0.28	0.13	0.13	64	0.26	0.19	0.18
1	47	0.21	0.16	0.14	52	0.19	0.20	0.19

Notes. The table records the standard deviations across individual respondents' EAU and EPU measures. We also record the standard deviation of individuals' RMSEs. The number of respondents underlying each cross-sectional standard deviation is recorded in the table.

Table 3: Spearman rank order correlation coefficients between individual EAU and EPU estimates

$h =$	8	7	6	5	4	3	2	1
				Output growth				
(1) EAU and EPU. No weighting	0.13	0.12	-0.28	-0.12	-0.14	0.30	0.24	0.27
(2) EAU and EPU. Weighted	0.12	0.27	-0.32	0.00	0.02	0.37	0.08	-0.03
(3) EAU and RMSE. Weighted	0.14	0.28	-0.26	0.02	-0.08	0.37	0.13	-0.05
				Inflation				
(1) EAU and EPU. No weighting	-0.46	-0.12	-0.14	-0.11	-0.35	0.04	0.27	-0.01
(2) EAU and EPU. Weighted	-0.28	-0.36	-0.17	-0.51	-0.24	0.09	0.13	-0.19
(3) EAU and RMSE. Weighted	-0.28	-0.49	-0.17	-0.42	-0.08	-0.13	0.21	-0.15

Notes: (1) Reports the rank order correlation coefficient between the unweighted *ex ante* and *ex post* forecast standard deviations, (2) weights these as in equations (7) and (6), and (3) compares equations (7) and (5). For a two-sided test of the null of no monotonic relationship at the 10% level the critical values are ± 0.352, and at the 5% level ± 0.415. (There are 23 observations in each case.)

Table 4: Interval coverage rates

	Output growth						Inflation					
	Centred on mean			Centred on point			Centred on mean			Centred on point		
	none	b.c	#	none	b.c	#	none	b.c	#	none	b.c	#
	50% nominal coverage rate											
8	0.28	0.34	650	0.31	0.36	636	0.37	0.44	629	0.37	0.45	616
7	0.32	0.36	740	0.32	0.36	740	0.35	0.48	728	0.40	0.46	728
6	0.35	0.37	732	0.34	0.36	732	0.37	0.50	717	0.44	0.49	717
5	0.35	0.35	797	0.34	0.35	797	0.42	0.48	785	0.46	0.52	785
4	0.38	0.40	653	0.40	0.42	639	0.52	0.54	634	0.57	0.60	621
3	0.54	0.54	745	0.59	0.61	745	0.56	0.64	730	0.71	0.76	730
2	0.59	0.60	737	0.63	0.61	737	0.58	0.62	723	0.79	0.81	723
1	0.53	0.50	798	0.78	0.73	798	0.53	0.54	783	0.90	0.84	783
	90% nominal coverage rate											
8	0.60	0.66	650	0.61	0.65	636	0.72	0.79	629	0.71	0.78	616
7	0.62	0.67	740	0.63	0.68	740	0.73	0.81	728	0.76	0.82	728
6	0.63	0.67	732	0.65	0.67	732	0.75	0.83	717	0.78	0.85	717
5	0.64	0.67	797	0.64	0.68	797	0.79	0.82	785	0.80	0.82	785
4	0.74	0.78	653	0.77	0.80	639	0.88	0.88	634	0.88	0.90	621
3	0.88	0.89	745	0.90	0.91	745	0.88	0.93	730	0.95	0.96	730
2	0.89	0.91	737	0.91	0.93	737	0.90	0.92	723	0.97	0.96	723
1	0.81	0.81	798	0.95	0.94	798	0.86	0.86	783	0.96	0.95	783

Notes: The prediction intervals are calculated by fitting normal distributions to the histograms ('Centred on mean: none'); by bias-correcting the histogram means on which the intervals are centred ('Centred on mean: b.c'); by centring the intervals on the point predictions ('Centred on point: none'); by bias-correcting the point predictions on which the intervals are centred ('Centred on point: b.c'). # denotes the number of forecasts (across individuals and time periods for a given h). Coverage rates are calculated uisng real-time second-release actuals.

Table 5: Summary of tests of individuals - Proportion of regressions for which we reject $E\left(w_{i,t|t-h}\right)^2 = 1$

			1-sided, $H_a < 1$, α% level				1-sided, $H_a > 1$, α% level				2-sided, α% level			
α	No.	No.	Unadjust.		Bias adjust.		Unadjust.		Bias adjust.		Unadjust.		Bias adjust.	
	regns.	obs	1-4	5-8	1-4	5-8	1-4	5-8	1-4	5-8	1-4	5-8	1-4	5-8
Output growth														
5	107	14	0.28	0.01	0.31	0.02	0.07	0.41	0.07	0.39	0.28	0.21	0.27	0.27
10	107	14	0.34	0.01	0.36	0.03	0.16	0.68	0.15	0.68	0.36	0.42	0.38	0.41
Inflation														
5	98	14	0.62	0.10	0.65	0.18	0.02	0.29	0.02	0.23	0.57	0.31	0.64	0.30
10	98	14	0.63	0.10	0.67	0.22	0.03	0.39	0.04	0.32	0.64	0.39	0.67	0.42

For a given forecast horizon, for each individual with a sufficient nuber of forecast observations, we regress either $w^2_{i,t|t-h}$ ('Unadjusted') or $[(e_{i,t|t-h} - \overline{e}_{i,h})/\sigma_{i,t|t-h}]^2$ ('Bias adjusted'), on a constant, and test the hypothesis that the constant is one. The table reports the proportion of regressions for which we reject the null against: i) a one-sided alternative that the constant is less than one (corresponding to 'underconfidence'); a one-sided alternative that the constant is greater than one (corresponding to 'overconfidence'); and iii) a two-sided alternative. We report rejection rates for two significance levels, α. We consider together all the within-year forecasts (denoted '1-4') and all the next-year forecasts (denoted 5-8).

The first three columns report the significance level, the number of regressions, and the average number of forecast observations in the regressions.

Table 6: Consensus forecasts: Subjective and *Ex Post* Inflation and Output Growth Forecast Uncertainty, Excluding surveys from 2007-.

h	Annual Output Growth								Annual Inflation							
	Subjective (EA)				*Ex Post*		*Ex Post*		Subjective (EA)				*Ex Post*		*Ex Post*	
	σ_h^m	s.	σ_h^p	s.	$\widehat{\sigma}_{h,ep}$	s.	RMS	s.	σ_h^m	s.	σ_h^p	s.	$\widehat{\sigma}_{h,ep}$	s.	RMS	s.
8	1.12	1.00	1.03	1.00	1.21	1.00	1.26	1.00	0.91	1.00	0.86	1.00	0.79	1.00	0.87	1.00
7	1.07	0.96	1.00	0.97	1.29	1.06	1.30	1.03	0.92	1.02	0.89	1.03	0.67	0.86	0.82	0.94
6	1.04	0.93	0.96	0.93	1.18	0.97	1.19	0.94	0.87	0.96	0.85	0.99	0.59	0.76	0.72	0.83
5	0.97	0.87	0.89	0.87	1.05	0.87	1.10	0.87	0.81	0.90	0.79	0.92	0.51	0.65	0.56	0.64
4	0.92	0.82	0.84	0.81	0.79	0.65	0.90	0.71	0.79	0.88	0.74	0.86	0.38	0.49	0.38	0.44
3	0.84	0.75	0.77	0.74	0.54	0.45	0.54	0.43	0.75	0.83	0.71	0.83	0.27	0.34	0.30	0.34
2	0.71	0.63	0.64	0.62	0.39	0.32	0.39	0.31	0.67	0.74	0.62	0.72	0.22	0.28	0.23	0.27
1	0.56	0.50	0.47	0.45	0.30	0.24	0.30	0.23	0.54	0.60	0.49	0.57	0.19	0.25	0.19	0.22

Notes: As table 1.

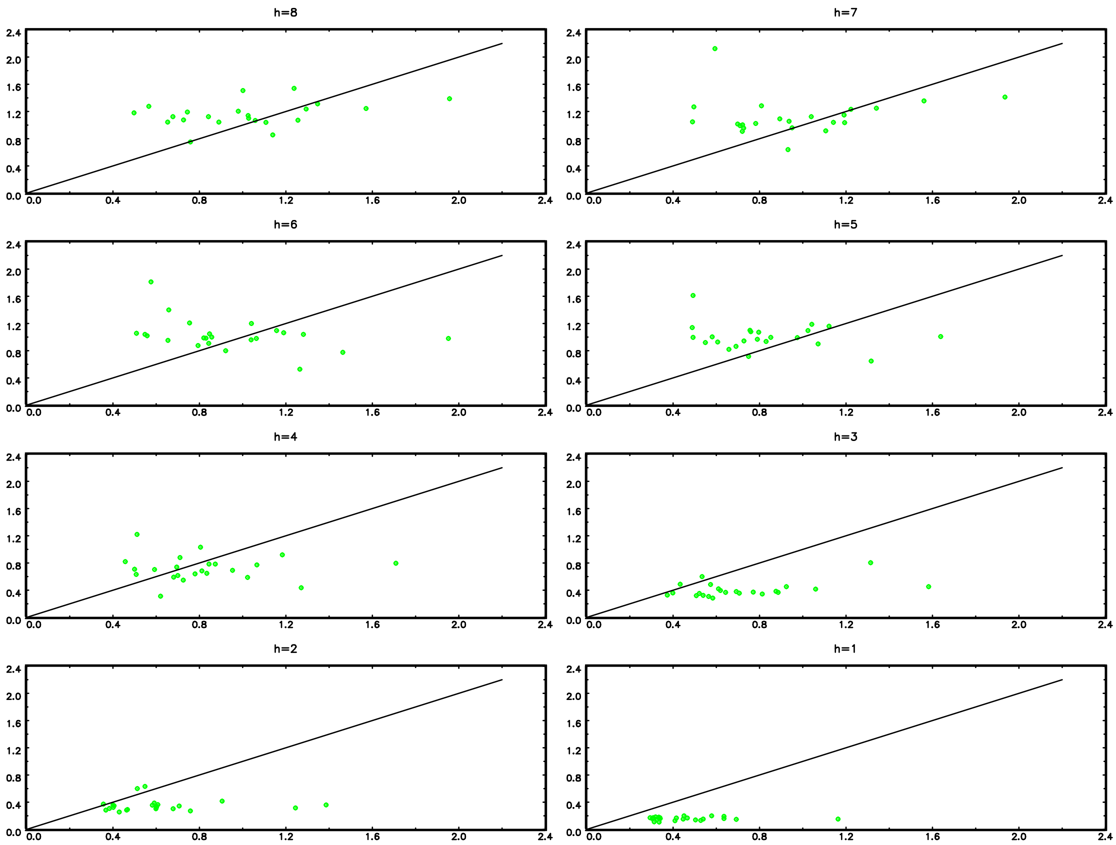

Figure 1: Output growth. Scatter plot of individual respondents' weighted estimates of EAU and RMSE. (EAU is plotted on the x-axis, and RMSE on the y-axis).

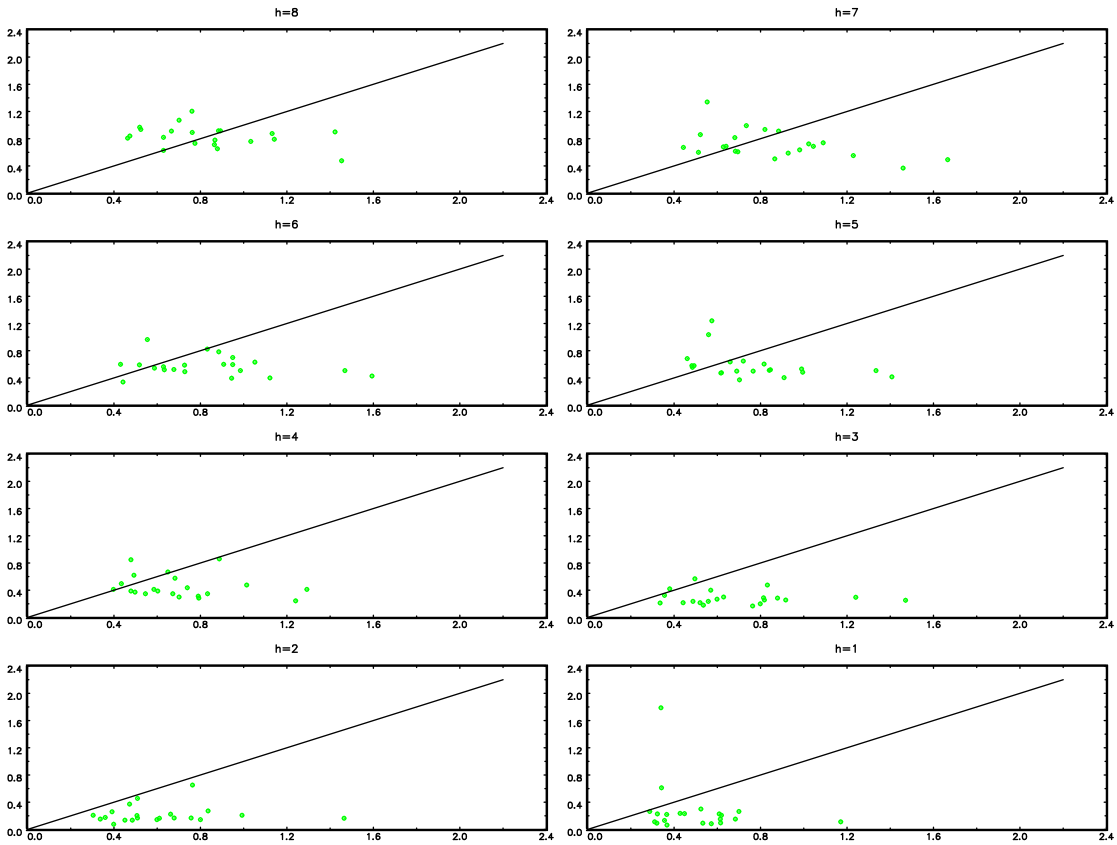

Figure 2: Inflation. Scatter plot of individual respondents' weighted estimates of EAU and RMSE. (EAU is plotted on the x-axis, and RMSE on the y-axis).

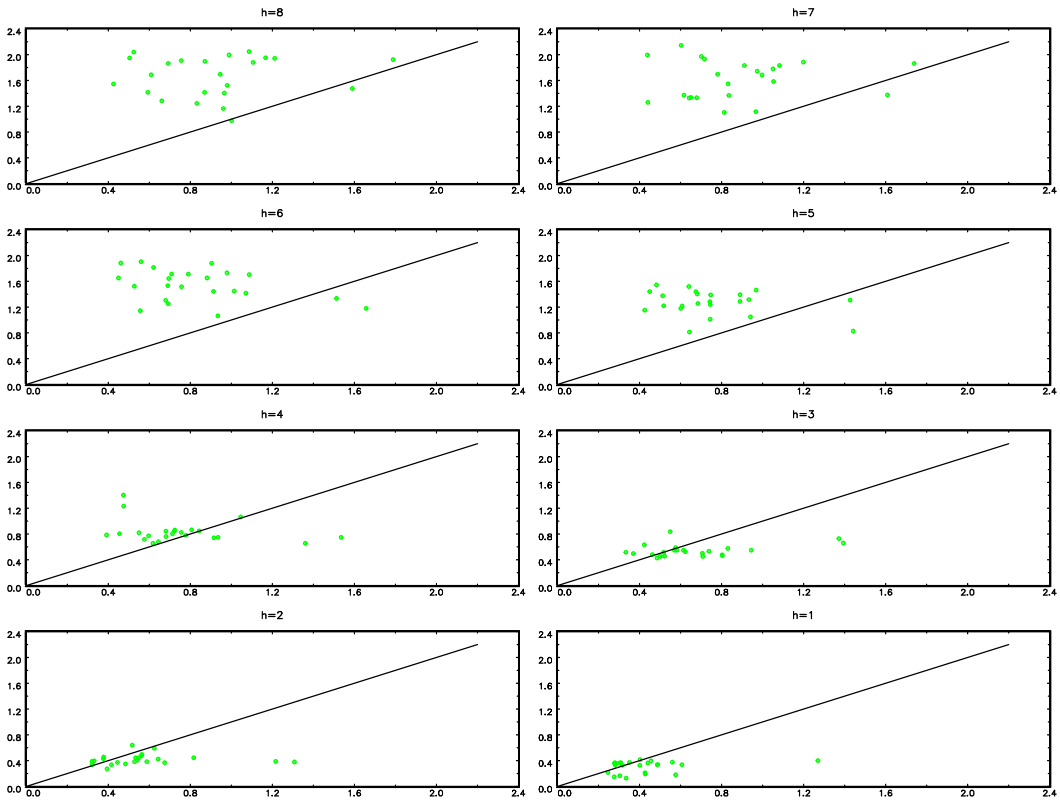

Figure 3: Output growth. Scatter plot of individual respondents' estimates of EAU and EPU. (EAU is plotted on the x-axis, and EPU on the y-axis).

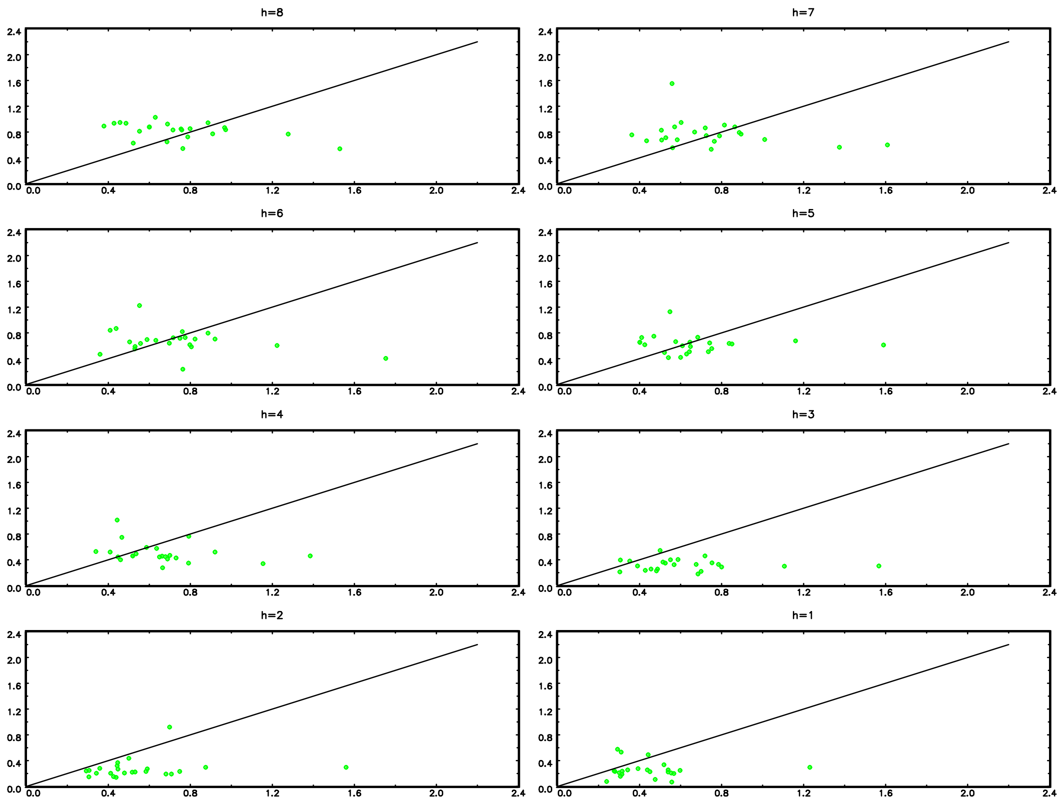

Figure 4: Inflation. Scatter plot of individual respondents' estimates of EAU and EPU. (EAU is plotted on the x-axis, and EPU on the y-axis).