Are Professional Forecasters Bayesian?

SEBASTIANO MANZAN

Bert W. Wasserman Department of Economics & Finance

Zicklin School of Business, Baruch College, CUNY

55 Lexington Avenue, New York, NY 10010

phone: +1-646-312-3408, email: *sebastiano.manzan@baruch.cuny.edu*

Abstract

I investigate how professional forecasters update their uncertainty forecasts of output and inflation in response to macroeconomic news. I obtain a measure of individual uncertainty from the density forecasts of the Survey of Professional Forecasters for the United States (US-SPF) and the Euro area (ECB-SPF) and use it to test the prediction of Bayesian learning that uncertainty should decline as the forecast date nears the target date. Empirically, I find that the prediction is occasionally violated, in particular when forecasters experience unexpected news in the most recent data release, and following quarters in which they produce narrow density forecasts. The evidence indicates also significant heterogeneity in the updating behavior of forecasters in response to changes in these variables. In addition, I propose a method to solve the problem of the truncation of the density forecasts that occurs when a significant amount of probability is assigned to the open intervals.

JEL Classification: E17, E37

Keywords: Bayesian learning, density forecasts, Survey of Professional Forecasters

Acknowledgments: I would like to thank the Associate Editor and two referees for many insightful comments as well as seminar participants at De Nederlandsche Bank, University of York, Queens College, the 2014 Conference in Real-Time Data Analysis, Methods, and Applications, and the 22nd Computing in Economics and Finance conference. Financial support by the PSC-CUNY Research Award Program and the Bert W. Wasserman fund is gratefully acknowledged. The usual disclaimers apply.

1 Introduction

The analysis of survey expectations shows that consumers and professional forecasters have diverging views about the future evolution of economic variables, which calls for a better understanding of the mechanism that agents use to form and revise their expectations (Mankiw *et al.*, 2004, and Dovern *et al.*, 2012). Mankiw and Reis (2002) propose a theory of sticky information in which agents update their forecasts only occasionally due to the cost involved in processing the newly released information. This produces dispersion in forecasts since at any point in time there is co-existence of agents that incorporate the most recent macroeconomic information while others persist using outdated forecasts. An alternative argument for the existence of heterogeneous beliefs among agents is that they update their forecasts at every point in time, but are limited in their ability to process public information (Woodford, 2002, and Sims, 2003). Andrade and Le Bihan (2013) and Coibion and Gorodnichenko (2015) provide empirical evidence that support the relevance of models with information rigidities based on survey expectations. Another argument for the existence of heterogeneous expectations is that agents use different models to form their expectations (Kandel and Pearson, 1995, and Brock and Hommes, 1997, 1998). Agents might produce different forecasts because they hold diverging prior views but also because, despite common priors, they interpret differently the relevance of the newly released information. There are several recent papers that try to disentangle these effects based on survey expectations. Lahiri and Sheng (2008) found that belief heterogeneity is largely due to differences in priors at long forecast horizons while it is driven by differential interpretation of news at short horizons. Using the same survey data but a different modeling strategy, Patton and Timmermann (2010) confirm that differences in priors represent the most important source of heterogeneity, although their results point to a minor role for the diversity in the interpretation of the signal. On the other hand, Manzan (2011) abstracts from the role of prior expectations to focus on the interpretation of news and finds evidence that forecasters are significantly heterogeneous in the way they update their forecasts. Overall, these papers suggest that forecasters are different in the way they form their prior expectations and in the way that they interpret the signal, although it is empirically difficult to disentangle the different effects without imposing any modeling assumption.

The aim of this paper is to investigate empirically what can be learned from *density* forecasts, rather than *point* forecasts, about the expectation formation process and the heterogeneity among forecasters. The density forecasts are obtained from the Survey of Professional Forecasters for the United States (US-SPF) and the Euro area (ECB-SPF). These Surveys collect expectations about the distribution of output growth and inflation by professional forecasters, in addition to point forecasts that have been used in some of the studies discussed earlier. The Surveys require participants to provide density forecasts from the first to the fourth quarter of a year with the goal of predicting the growth rate of the variable for that year. This structure thus allows to focus the analysis on the effect that macroeconomic news have on the revision of density forecasts since the forecast target remains constant while the horizon shortens over time. The empirical analysis is conducted adopting a simple Bayesian Learning Model (BLM) as the framework for the formation and revision of expectations. The model predicts that the posterior mean equals the weighted average of the prior mean and the signal contained in the new data releases, while the posterior precision (inverse of the variance) is given by the sum of the prior and the signal precisions. In particular, in this paper I test empirically the validity of the BLM prediction that the precision of density forecasts should not decline as the forecast horizon shortens. A forecaster that believes that the incoming data is uninformative about the future outcome of the variable will assign zero precision to the signal and thus expect that the posterior and prior precisions are equal. On the other hand, a forecaster that believes in the informational content of the data release will assign higher precision to the posterior, relative to the prior.

Our empirical strategy consists of using the density forecasts to construct an observable measure of uncertainty for each forecaster at each point in time. This is done by calculating the variance of the individual densities which I then use to calculate the ratio of the prior and posterior precisions. The variance can then be used to evaluate if forecasts produced by professional forecasters are consistent with Bayesian updating of the precision and to what extent they are different across forecasters. The empirical evidence shows frequent violations of the BLM prediction since forecasters often provide density forecasts for output growth and inflation that are more uncertain (less precise) relative to the forecasts they produced in the previous quarter. This is inconsistent with Bayesian learning since the additional macroeconomic data released in the current quarter should increase their expected precision in case they consider the data informative or it should not change the prior precision in case they regard the news as uninformative. In addition, I find that the non-Bayesian behavior is common among most professional forecasters in the sample, although the frequency and magnitude of the deviations might be different.

The empirical analysis shows three main factors driving these findings. The first is a methodological issue with the administration of the Surveys. Forecasters are asked to assign probabilities to a set of intervals including two open intervals at the extremes of the grid. In most quarters, forecasters assign probabilities to the inner intervals of the grid, although occasionally they have assigned probability to the open intervals. An instance of this is the first quarter of 2009 when the rapid deterioration of macroeconomic conditions in Europe and the rest of the world caused many ECB-SPF forecasters to assign high probability for GDP growth to the lowest open interval of -1% or lower. In the following quarter, the left open interval was shifted to -6% or lower that allowed forecasters to assign most probability to the inner intervals. The truncation of the density forecasts in the first quarter of 2009 constitutes a problem when the analysis relies on extracting the mean and variance from the density forecasts. I solve this problem by constructing pseudo-density forecasts by exploiting the fact that forecasters are asked point forecasts, in addition to density forecasts. I assume that the point forecast represents the approximate center of the pseudo-density forecast and that the underlying distribution is triangular. In this way, I am able to produce pseudo-histograms that I interpret as the distribution that the forecaster would have submitted had the truncation not occurred.

Once this methodological issue is resolved, I find that individuals are likely to reduce the precision of their density forecasts, relative to their prior quarter forecast, when they receive a large surprise and as a reaction to having produced a very narrow distribution in the previous quarter. I define the surprise as the most recent data release of the variable being forecast, standardized with the mean and standard deviation of the individual prior density forecast. The evidence indicates that both negative and positive surprises have the effect of increasing the ratio of the prior to posterior precision. This suggests that, at times, forecasters react to unexpected news by increasing the dispersion of their density forecasts, relative to their prior expectation. The second factor that I find relevant in causing the non-Bayesian updating is the fact that forecasters sometimes concentrate their density forecast in a few bins. Due to the discreteness imposed by the histogram, a shift in the mean of a density that is highly concentrated in a few bins might cause the probability to be more disperse relative to the previous quarter. I find that the smaller the number of bins used in a quarter the more likely it is that in the following quarter the forecaster decreases the precision of the density. Finally, by using a panel group estimator we are able to identify groups of forecasters with an homogeneous response to surprises within the group, but heterogeneous across groups.

This paper is organized as follows. In Section (2) I introduce the BLM and discuss its implications for the updating behavior of professional forecasters. In Section (3) I discuss the density forecasts provided by the US-SPF and ECB-SPF and in Section (4) I conduct an exploratory analysis of the empirical support for the BLM prediction followed by a regression analysis to understand the determinants of the observed non-Bayesian behavior of forecasters. Finally, Section (6) draws the conclusions of the paper.

2 Bayesian Learning Model

I discuss the Bayesian Learning Model (BLM) in the context of the information arrival and expectation formation of the US-SPF[1]. In quarter 1 of year t forecasters observe the first release of real GDP and GDP deflator for quarter 4 of year $t-1$. This allows them to calculate the average level of the variable for the previous year, which is given by $\bar{Y}_{t-1} = \sum_{q=1}^{4} Y_{q,t-1}/4$, where $Y_{q,t-1}$ denotes the level of real GDP or the GDP deflator in quarter q of year $t-1$. After observing the first release for quarter 4 of the previous year, the forecaster is asked to provide an expectation about the (year-over-year) growth rate of the variable in year t which I indicate by y_t and is defined as $y_t = (\bar{Y}_t - \bar{Y}_{t-1})/\bar{Y}_{t-1}$. I denote the quarter 1 density forecast of an agent by $f_1(y_t)$, where the subscript 1 indicates the quarter the forecast is made. Notice that as forecasters form an expectation for y_t in the first quarter they only observe the realization of the variable for the previous quarter due to the lag in releasing economic data by the statistical agencies. In quarter q (for $q = 2, 3, 4$) the forecaster observes the first data release for the previous quarter of that year, $Y_{q-1,t}$, and provides a density forecast, denoted by $f_q(y_t)$. The year-over-year growth rate y_t can be decomposed in a component that has already been released and another that is unknown since it involves the current and future

[1]The only difference in the ECB-SPF is the GDP release that arrives with a two quarter publication lag, while the price indicator is available shortly after the end of the month.

Quarter	$\bar{y}_{q,t}$	$\tilde{y}_{q,t}$	$E_q(\tilde{y}_{q,t})$	$Var_q(\tilde{y}_{q,t})$
2	$\frac{Y_{1,t}}{4\bar{Y}_{t-1}}$	$\frac{Y_{2,t}+Y_{3,t}+Y_{4,t}}{4\bar{Y}_{t-1}}$	$3l_{2,t}$	$(3+4\theta_{2,t})\sigma^2_{2,t}$
3	$\frac{Y_{1,t}+Y_{2,t}}{4\bar{Y}_{t-1}}$	$\frac{Y_{3,t}+Y_{4,t}}{4\bar{Y}_{t-1}}$	$2l_{3,t}$	$2(1+\theta_{3,t})\sigma^2_{3,t}$
4	$\frac{Y_{1,t}+Y_{2,t}+Y_{3,t}}{4\bar{Y}_{t-1}}$	$\frac{Y_{4,t}}{4\bar{Y}_{t-1}}$	$l_{4,t}$	$\sigma^2_{4,t}$

Table 1: The year-over-year growth rate of Real GDP and PGDP y_t is decomposed as in Equation (1) in a part that is known in quarter q, denoted by $\bar{y}_{q,t}$, and another that is still unknown and denoted by $\tilde{y}_{q,t}$. The following columns provide the mean and variance expectations for the year-over-year growth rate y_t based on the assumption discussed below.

quarters. In quarter q $(= 2, 3, 4)$ the growth rate can be expressed as

$$y_t = \sum_{k=1}^{q-1} \frac{Y_{k,t}}{4\bar{Y}_{t-1}} + \sum_{j=q}^{4} \frac{Y_{j,t}}{4\bar{Y}_{t-1}} - 1 = \bar{y}_{q,t} + \tilde{y}_{q,t} - 1 \tag{1}$$

where $\bar{y}_{q,t}$ represents the portion of the annual growth rate y_t that can be calculated based on the released data up to quarter q, while $\tilde{y}_{q,t}$ represents the current and future quarterly growth rates that will become available in the following quarters. The values of $\bar{y}_{q,t}$ and $\tilde{y}_{q,t}$ in quarters 2 through 4 are provided in Table (1). In the second quarter forecasters only know the data release for the first quarter, $Y_{1,t}$, and have to form expectations about the current and future quarters. On the other hand, in quarter 4 they know the realizations for the first three quarters and the only uncertainty in forecasting y_t derives from the quarter 4 realization, $Y_{4,t}$, which will be released in quarter 1 of year $t + 1$. Given the timing of information arrival discussed above, forecasters update their density expectations for y_t in quarter 2, 3, and 4. I assume that individuals form an expectation about the quarterly change of the variable, denoted by $y_{q,t} = Y_{q,t}/(4\bar{Y}_{t-1})$, and that they expect the forecast to hold also for future quarters. I also conjecture that agents interpret the new information as a signal for $Y_{q,t}$ ($q = 2, 3,$ and 4) which has mean $L_{q,t}$ and variance $\sigma^2_{q,t}$. In addition, I assume the forecaster believes that the expectation of the stadardized quarterly change $y_{q,t} = Y_{q,t}/(4\bar{Y}_{t-1})$ is given by $E(y_{q,t}) = L_{q,t}/(4\bar{Y}_{t-1}) = l_{q,t}\ \forall q$ and that the quarterly changes of the variable are first-order correlated, i.e., $corr(y_{q,t}, y_{q+1,t}) = \theta_{q,t}$. Based on these assumptions, the forecaster interprets the recently released data as a signal about $\tilde{y}_{q,t}$ which I assume is normally distributed with mean $E_q(\tilde{y}_{q,t})$ and variance $Var_q(\tilde{y}_{q,t})$ and the values for each q are provided in Table (1). This allows to calculate the mean expectation for the year-over-year growth rate y_t given by $E_q(y_t) = \bar{y}_{q,t} + E_q(\tilde{y}_{q,t}) - 1$ and precision $\phi_{q,t} = Var_q^{-1}(y_t) = Var_q^{-1}(\tilde{y}_{q,t})$.

After interpreting the signal, the agent revises the mean and precision of the density forecast. Since both the prior and signal are normally distributed and assumed to be independent, the posterior mean and precision are given by

$$\mu_{q,t} = \rho_{q,t}\mu_{q-1,t} + (1 - \rho_{q,t})E_q(y_t) \tag{2}$$

$$\psi_{q,t} = \psi_{q-1,t} + \phi_{q,t} \tag{3}$$

where $\rho_{q,t} = \psi_{q-1,t}/(\phi_{q,t} + \psi_{q-1,t})$ represents the weight assigned to the prior (relative to the signal) and is given by the ratio of the precision of the prior to the precision of the posterior distribution (given in Equation 3). The posterior mean in quarter q is the weighted average of the prior mean and the expected effect of the released information, with the weight assigned to each component depending on the subjective precision of the signal and prior. If the agent believes that the signal does not provide any insight on y_t he/she will expect the precision $\phi_{q,t}$ to be equal to zero so that $\rho_{q,t}$ will take a value of 1. On the other hand, if the forecaster believes that the signal is very informative about y_t then the signal will be expected to have high precision and thus $\rho_{q,t}$ will be close to 0 so that the posterior mean will be close to the signal.

The BLM represents an expectation formation model that provides restrictions on the revisions of expectations that can be tested empirically. Kandel and Pearson (1995) use this model to evaluate the revisions of earning forecasts around news releases and find evidence that a significant fraction of analysts revise their forecasts in a manner that is inconsistent with the common interpretation of the signal. There are also several applications of the BLM to macroeconomic expectations, such as Kandel and Zilberfarb (1999), Lahiri and Sheng (2008) and Patton and Timmermann (2010). Manzan (2011) estimates the Bayesian weight coefficients $\rho_{q,t}$ based on point predictions and assuming that the signal is represented by the latest release for the variable being forecast. He finds significant evidence of weight heterogeneity across forecasters at most horizons. These studies use only point forecasts to test their hypothesis about the learning model. Instead, in this paper I propose to look at the second moment of the subjective distribution forecasts to test hypothesis about the expectation formation process. In particular, the model implies that as time advances from the first to the fourth quarter, the individual precision of forecaster i in quarter q is given by

$$\psi_{q,t} = \psi_{1,t} + \sum_{j=2}^{q} \phi_{j,t} \tag{4}$$

for $q = \{2, 3, 4\}$. This Equation shows that the posterior precision of a forecaster is the sum of the prior precision in the first quarter and the cumulative precision of the signals. This implies that the posterior precision $\psi_{q,t}$ should not decrease as the target date gets closer since the $\phi_{q,t}$ are necessarily non-negative. This is a sensible prediction since uncertainty in forecasting the same quantity should reduce closer to the target date. Hence, for a forecaster that updates in a Bayesian manner it should hold that $\psi_{q,t} \geq \psi_{q-1,t}$, with equality holding only when the forecaster believes the latest signal is totally uninformative. A related prediction that emerges from the BLM model concerns the Bayesian weights $\rho_{q,t}$. Since the weight is the ratio of the prior and the posterior precision, the restriction in Equation (4) implies that

$$0 \leq \rho_{q,t} = \frac{\psi_{q-1,t}}{\psi_{q,t}} \leq 1 \tag{5}$$

The weight $\rho_{q,t}$ is thus bounded to be smaller or equal to 1 with the constraint binding when the forecaster assigns zero precision to the signal in the current quarter. In addition, the weight is also restricted to be non-negative since it represents the ratio of two precisions (or variances).

The previous discussion suggests that forecasters updating their expectations in a Bayesian manner should be characterized by non-decreasing precisions of their posterior density as the target date approaches, which also implies that the Bayesian weight used in revising their expectations should not be larger than 1. In other words, the arrival of new information is likely to shift the center of the density forecasts and reduce its dispersion closer to the target date. However, since the work of Tvesky and Kanheman (1974) there is growing evidence that agents deviate from Bayesian updating in laboratory experiments. Recently, Epstein (2006) and Epstein *et al.* (2008, 2010) have proposed theoretical models that account for this behavior. They argue that non-Bayesian updating might arise in response to a signal that is interpreted as positive or negative by forecasters. This might lead them to produce a posterior forecast that is inconsistent with their prior since, after the signal is observed, the forecaster believes in a different prior. The forecaster is thus updating in a Bayesian manner a prior which is different from the prior he/she expected before the signal was observed. Ortoleva (2012) suggests that agents might re-assess their prior forecasts in response to signals that have low probability to occur based on their prior belief. In this case, a forecaster interprets the low probability event as an indication of misspecification of the forecasting model which prompts the revision of her prior belief. In a similar spirit, Nimark (2014) proposes a model in which an unusual event increases the probability of observing a public signal. The signal leads agents to adopt a prior forecasts that has higher dispersion relative to the prior they would update if no signal was observed. This might lead to the precision of the posterior forecast in quarter $t+1$ to be smaller relative to prior at time t because, after the public signal is observed, the agent believes in a different prior.

Although these models motivate differently the non-Bayesian updating, they share a common mechanism that relates the interpretation of news to a revision of the prior belief. If this is the case, then we might observe quarters in which some forecasters report a posterior precision $\psi_{q,t}$ that is smaller relative to the prior precision $\psi_{q-1,t}$ and the weight $\rho_{q,t}$ that can take values larger than one. In Section (4) I obtain the precision for each forecaster-quarter for the US-SPF and the ECB-SPF and evaluate the hypothesis that forecasters update in a Bayesian manner their prior expectations.

3 Survey Density Forecasts

To empirically test the hypothesis of Bayesian updating discussed above, I consider two Surveys that ask respondents to provide density forecasts of output and inflation. The Surveys are the ASA-NBER-Philadelphia Fed Survey of Professional Forecasters (US-SPF) and the European Central Bank Survey of Professional Forecasters (ECB-SPF) for the Euro area. The Surveys ask professional forecasters employed in the private sector and research organization to provide point and density forecasts of output and inflation at a horizon ranging from the current year up to 2-3 years ahead. The output variable for both Surveys is real GDP and the inflation measure is the GDP deflator for the US and the Harmonised Index of Consumer Prices (HICP) for Europe. The US-SPF started in 1968 as a joint initiative of the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER) and since 1992 it is administered by the Federal Reserve Bank of Philadelphia (see Croushore *et al.*, 2019, for more details and a recent assessment of 50 years of the Survey). Instead, the European Central Bank (ECB) administers the ECB-SPF since the first quarter of 1999 (see Garcia, 2003).

In both Surveys, forecasters are asked to predict the probability that the growth rate of the variables will fall in specified intervals, with the first and last being open intervals. One difference between the Surveys is that the US-SPF requires the forecast of the average-over-average growth rate of the variable while the ECB-SPF asks for the growth rate between quarter 4 of the current year relative to quarter 4 of the previous year. An interesting feature of the ECB-SPF is that it provides both fixed and changing horizon point and density forecasts. Fixed horizon forecasts are expectations about the value or distribution of the variable at a constant horizon h, as opposed to changing horizon in which the forecast target date (e.g., year t or $t + 1$) is constant. In this paper we only consider the changing horizon (also called fixed target) forecasts from the ECB-SPF since they are consistent with the survey scheme used for the US-SPF and the Bayesian learning model discussed above.

The density forecasts from these Surveys are unique among expectations data because they provide a measure of the mean/median outcome expected by forecasters, but also an individual measure of the expected uncertainty about the future growth in output and prices. Several studies have investigated the properties of these density forecasts, in particular for the US-SPF. Zarnowitz and Lambros (1987) is one of the first analysis of the characteristics and properties of density forecasts. An issue that they investigate is the relationship between the cross-sectional dispersion of point forecasts (i.e., disagreement) and aggregate measures of uncertainty. They measure uncertainty by the standard deviation of the consensus density obtained by averaging the individual density forecasts. They found that disagreement and uncertainty are positive correlated which provides support for the use of measures of forecast disagreement as an observable proxy for macroeconomic uncertainty. This conclusion has later been revisited using longer sample periods, individual rather aggregate density forecasts, and alternative measures of uncertainty derived from the histogram forecasts. Giordani and Söderlind (2003) find even stronger evidence to support these earlier findings, while Lahiri and Liu (2006) and Rich and Tracy (2010) conclude that there is weak evidence of a relationship between disagreement and uncertainty using alternative measures of uncertainty. Another issue that has received attention in the literature is the mechanism used by forecasters to form expectations about future uncertainty. Giordani and Söderlind (2003) and Lahiri and Liu (2006) use GARCH-type specifications for the (observable) variance of the density forecasts and found that there is significant persistence in these measures, although the persistence is smaller than values obtained from aggregate time series. Lahiri and Liu (2006) also found that uncertainty is more responsive to expected increases of the inflation rate than to expected declines in the rate. Clements (2014) compares measures of forecast uncertainty with suitably constructed measures of ex-post uncertainty. He finds that forecasters seem to under-estimate dispersion at long horizons but over-estimate it at short forecast horizons when considering inflation and output growth. In addition, Sheng and Yang (2013) test the hypothesis that the precision in Equation (4) has a unit root and interpret the rejection for several forecasters as evidence that they update their density in a non-Bayesian way. Some other papers (see Engelberg *et al.*, 2009, Clements, 2010, and Abel *et al.*, 2016) have looked at the consistency of point and density forecasts produced by the same individuals. The findings indicate that point forecasts are closely related to measures of central tendency obtained from the density forecasts. More recently, Del Negro *et al.* (2018) propose a non-parametric approach to estimate the density forecast based on the histograms while Ganics *et al.* (2020) develop a method that converts the fixed-event density forecasts of the US-SPF in fixed-horizon forecasts.

Despite the short history of the ECB-SPF, there have been several studies that have investigated the properties of the forecasts, mainly in comparison to the results available for the US-SPF. An early paper is Bowles *et al.* (2007) that finds that forecasts of GDP and unemployment are not biased, although those for inflation consistently underpredict the realization. However, long-term forecasts of inflation seem to be well anchored to the ECB policy rate. Andrade and Le Bihan (2013) use the ECB-SPF to assess the hypothesis of sticky information by evaluating of adjustment of expectations. They find that forecasters adjust their expectations periodically rather than continuously and that there is significant disagreement in the way they update their forecast. The density forecasts have been analyzed by Abel *et al.* (2016) who find a weak relationship between uncertainty and forecast dispersion and Kenny *et al.* (2014, 2015) that evaluate and model the accuracy of individual density forecasts and find that accuracy of the density forecasts is negatively affected by a downward bias in the forecast of uncertainty.

3.1 Moments of the histogram

The Surveys provide each quarter a histogram representing the percentage probabilities assigned by a forecaster that the realization of the variable will fall in a certain interval. There are two approaches that have been used to extract moments from the histograms. D'Amico and Orphanides (2008) is a recent example of a distribution-free approach since they calculate the mean and variance of the distribution only based on the reported probabilities. Denote by $\bar{x}_j$ the mid-point of the j-th interval (for $j = 1, \cdots, J$) and by $p_{j,q,t}$ the probability assigned by a forecaster to interval j in quarter q of year t. Then, the mean and variance are calculated as follows:

$$\mu_{q,t} = \sum_{j=1}^{J} \bar{x}_j p_{j,q,t} \tag{6}$$

$$\sigma^2_{q,t} = \sum_{j=1}^{J} (\bar{x}_j - \mu_{q,t})^2 p_{j,q,t} - w^2/12 \tag{7}$$

where w is the bin width and the term $w^2/12$ represents the Sheppard's correction (Kendall and Stuart, 1977). The first and last intervals are open and I follow the conventional approach of interpreting the open interval as a close interval of the same size of the rest of the grid. An alternative approach is to fit a parametric distribution to the histograms, e.g., the normal as proposed by Giordani and Söderlind (2003). However, Engelberg *et al.* (2009) argue that the normal distribution might not be an appropriate assumption for the US-SPF density forecasts since it does not account for the evidence that forecasters often provide asymmetric histograms. As an alternative they propose to use the beta distribution defined on a finite interval which is able to account for the possible asymmetry of the density forecasts. Garcia and Manzanares (2007) provides a comparison of different distributional models including a skewed version of the normal distribution for the ECB-SPF. In the empirical application I will use the distribution-free approach to calculate the mean and the variance of the distribution only since the results based on the parametric approaches lead to very similar conclusions.

4 Descriptive Analysis

To empirically evaluate the prediction of the BLM I consider the current year forecasts from the US-SPF and ECB-SPF and include in the sample only those forecasters that provide at least 30 consecutive predictions over the period considered. For the ECB-SPF I use data from the beginning of the Survey in the first quarter of 1999 until the fourth quarter of 2013. Instead, for the US-SPF I use the sample period from the first quarter of 1992 until the fourth quarter of 2013. Although the US-SPF density forecasts are available since 1968, there are several issues in using forecasts produced before 1992 when the Philadelphia Fed started administering the Survey. First, the definition of the output measure changed from nominal to real in 1981 and from GNP to GDP in 1992, which makes the analysis over time challenging. Another concern is that the bin size of the survey histogram was changed from 1 to 2% during the 1980s for both GDP and the GDP deflator and was then changed back to 1% intervals in 1992. Lahiri and Wang (2020) show that the effect of wider bins is to overestimate uncertainty extracted from the histogram relative to the case with smaller bin sizes. After 1992 the bin size is set at 1% except starting in 2014Q1 for PGDP for which the interval size is set at 0.5%. We follow the suggestion of Lahiri and Wang (2020) and combine adjacent 0.5% intervals to 1% intervals, thus providing a consistent bin size over the full sample considered in this paper. In addition, the participation rate during the 1980s dropped dramatically until the Philadelphia Fed started administering the Survey. For these reasons I decided to start the Survey in 1992 which has also the advantage of having both Surveys spanning a similar period of time. The US-SPF panel consists of 22 forecasters while for the ECB-SPF there are 31 individuals contributing to the Survey.

4.1 Precision of the density forecasts

Figure (1) shows the quarterly time average of the precision of the density forecasts for the US-SPF and ECB-SPF for each forecaster in my panel. Each line is obtained by averaging the inverse of the variance calculated using the distribution-free approach discussed above. The four plots show a similar tendency of the average precision to increase toward the end of the year for most forecasters. However, there are notable differences in the evolution of the average precision among forecasters, horizons, variables and Surveys:

- In the case of the inflation measures the precision remains relatively constant in the first three quarters of the year and increases significantly in the last quarter. Instead, individuals seem to increase precision steadily throughout the year when forecasting real GDP. This suggests that a majority of forecasters interpret the quarterly macroeconomic releases as informative about output and decrease their subjective uncertainty about the current year realization of the variable.

- There is remarkable heterogeneity among forecasters in the level of precision at all horizons. In addition, the dispersion seems to become even more pronounced in the fourth quarter when forecasters know the realizations for the first three quarters and are only missing the quarter 4 release to be able to calculate the realization.

- Forecasters that are more (less) uncertain in the first quarter are also likely to be more (less) uncertain in the following quarters. I measure the persistence in the subjective belief of uncertainty by the rank correlation of the precision between pair of quarters. The evidence suggests that the correlations between current and following quarter precisions are between 0.69 and 0.81 in the first quarter and reduce to between 0.52 and 0.62 in quarter 3 across variables and Surveys. The rank correlations of the precision in quarter 1 and 4 ranges between 0.23 and 0.57. Overall, the evidence suggests that inflation and output uncertainty reduce during the year, but that the individual belief about uncertainty remains quite consistent over time with some forecasters expecting systematically higher uncertainty relative to other forecasters. The evidence of persistence in uncertainty has been documented by De Bruin *et al.* (2011) for consumer expectations and Boero *et al.* (2015) for density forecasts from the UK Survey of Professional Forecasters.

- Comparing the US-SPF and the ECB-SPF, it seems that forecasters are more confident about predicting the euro-area output and inflation relative to the US variables. The sample period is slightly different for the two Surveys, but the differences in precision are considerable and remain so even when compared for the same sample period.

- A close examination of the evolution of the precisions shows that for some forecasters the lines do not monotonically increase over time. This means that these forecasters expect (on average) more uncertainty (less precision) in the following quarters despite having observed the information contained in the recent data releases. This finding is inconsistent with the prediction of the BLM model that the precision of the density forecasts should be non-decreasing.

In Figure (1) I average the precision of each forecaster over time in order to gauge the characteristics of the density forecasts at the individual level. Instead, Figure (2) displays the time series evolution of the average precision in each quarter with the average calculated across forecasters. The plot also shows the recession periods indicated by the NBER for the US and by the CEPR for the Euro-area. Some facts emerge from these graphs:

- The time series for a certain quarter should lie above or equal the line for the previous quarter in order to be consistent with the prediction of the BLM. Although the prediction seems to hold most of the time, there are several instances in which two lines cross, which indicates that the cross-sectional average precision in a certain quarter is lower than the one in the previous quarter. An example of this situation is provided by the precision forecasts for real GDP growth for 2009 in Europe. For that year, the average precision for Q2, Q3 and Q4 are below the Q1 precision which means that in the second and following quarters of 2009 forecasters updated their density forecast by increasing their expected uncertainty.
- The average precisions by quarter vary significantly over time, in particular in Q4. In addition, the spread between the curves seems to be changing over time in response to economic events that might impact the agent expectations.
- There was a significant decrease in precision of European inflation in 2007 at all quarters and in the first three quarter for real GDP. This fact suggests that forecasters were more uncertain about the prospects of inflation and output which, in the case of inflation, persisted even in the fourth quarter. On the other hand, for the US I do not find evidence of a level-shift as it occurred for the euro-area.

The discussion so far has focused on the precision of the density forecasts produced by professional forecasters and we have considered the variation of the precision both across forecasters and over time. Another quantity that is interesting to consider is the precision of the signal defined as the difference between the prior and posterior precisions. The BLM model predicts that signal precision should be positive and in the following Section I empirically investigate its characteristics.

4.2 Precision of the signal

An alternative way to analyze the updating behavior of forecasters is to examine the precision of the signal that in the BLM is obtained as the difference between the posterior and prior precisions. The precision of the signal is non-negative and measures the confidence that a forecaster assigns to the recently released information. Large values of the signal precision indicate that the forecaster believes the news is highly informative about the realization of the macroeconomic variable. Figure (3) shows the variation over time of the average signal precision for the inflation and output measures in the two Surveys. In addition to significant variation over time of the precision, it appears that forecasters consider the information released in the last quarter as the most informative, in particular for output. Another fact that emerges from these plots is that in several quarters the precision of the signal is negative. In most cases the value is slightly negative but there are several episodes in which the precision is remarkably negative. In particular, this happened in the fourth quarter of 2003 for real GDP growth in the US and in the second quarter of 2009 for output growth in Europe. For the inflation series the quarters with negative precision are typically less extreme in magnitude.

What caused such dramatic revision of density forecasts by many forecasters in these quarters? In the United States, the Bureau of Economic Analysis (BEA) released on October 30th 2003 the advance estimate of real GDP growth for the third quarter of 2003 that reported an output increase of 7.2% at an annualized rate[2]. The news received a lot of media attention such as an article in the New York Times[3] entitled "*Economy records speediest growth since the mid-80s*" that argued: "*The economy expanded at the fastest rate since 1984 during the three months ended in September, the government reported yesterday, offering hope that the long economic malaise has finally ended. Consumer spending soared, foreigners bought American-made goods at a surprising clip and companies increased their investments in equipment and technology at a pace reminiscent of the 1990's boom.*" Figure (4) shows the consensus density forecast for the third and fourth quarters of 2003 calculated by averaging the probabilities in each bin. In quarter 3 the consensus density assigned 65% probability that the real output growth for 2003 would take a value between 2 and 3% and between 13-15% probability that the value will fall in the 1-2% and 3-4% bin. Following the BEA news release, forecasters revised their density forecast and the consensus shifted to the right with an increase in the probability that real GDP growth for 2003 would fall in the 3-4% and 4-5% intervals, while the bins between 0 and 3% were assigned smaller probabilities. The shift is also clear from the mean of the distribution that increases from 2.43% to 2.99% while the standard deviation decreases from 0.80% to 0.68%. The consensus forecast is thus consistent with the Bayesian prediction of lower uncertainty closer to the target date, although it hides the considerable heterogeneity in the expected uncertainty among forecasters. Considering the density forecasts at the individual level, I find that 13 of the 20 forecasters that participated to the Survey in both quarters reported higher uncertainty in their quarter 4 distribution forecast relative to quarter 3. This can be seen in Figure (5) that shows the histograms in quarter 3 and 4 of 2003 for the 20 forecasters along with the weight that represents the ratio of the prior and posterior precisions. The plots are sorted by the value of the Bayesian weight of each forecaster in the last quarter of 2003. All forecasters reacted to the news in quarter 4 by shifting probability to the right relative to the previous quarter density. This led to an increase of the dispersion of the distribution as in the case of the first four forecasters that assigned 100% probability to the 2-3% interval in the third quarter and later revised the distribution by assigning probability also to the 3-4% interval. A possible rationale for this behavior is that a forecaster might believe in Q3 that the distribution is uniform in the interval 2-3%. After the news is released, the forecaster shifts the distribution to be uniformly distributed in the interval 2.5-3.5%. Since the Survey requires to report the probability on the 2-3 and 3-4% intervals, the individual might thus report a 50%

[2] https://www.bea.gov/newsreleases/national/gdp/2003/gdp303a.htm

[3] http://www.nytimes.com/2003/10/31/business/economy-records-speediest-growth-since-the-mid-80-s.html

probability to each bin. Although it appears that uncertainty has increased, the forecaster did not change the prior views about uncertainty. Hence, the discretization of the histogram might also be responsible for some of the inconsistency that we have documented above.

The ECB-SPF Survey for the first quarter of 2009 was conducted between January 15 and 20[4] and the latest data available about real GDP was released by Eurostat on January 8, 2009[5]. The data released was the second estimate of GDP growth in the third quarter of 2008 and the announcement was that "*GDP declined by 0.2% in both the Euro area (EU15) and EU27 during the third quarter of 2008, compared to the previous quarter*". On January 15 2009 the Governing Council of the ECB decided to reduce key policy rates by 50 basis points to 2% in response to "*the latest economic data releases and survey information, which add clear further evidence to the assessment that the euro area economy is experiencing a significant slowdown, largely related to the effects of the intensification and broadening of the financial turmoil.*" The weakening of the European and world economies and the actions of the ECB contributed to shifting forecaster's expectations toward gloomier outcomes. The consensus density forecast experienced a dramatic shift toward more negative outcomes and also an increase in its dispersion. Figure (6)[6] shows a similar behavior for 20 individual density forecasts. However, further analysis indicates that part of the shift is due to the effect of the truncation of the grid: in the first quarter of 2009 the lowest bin in the Survey was an open interval for values lower than minus 1%. The large probabilities that are assigned to the interval between minus 1% and minus 1.5% reflect the conventional choice of interpreting the open interval as a closed interval with width 0.5% which is the bin size used in the ECB-SPF Survey. However, in the second quarter of 2009 the ECB changed the grid used in the Survey and shifted the left open interval to values equal or lower than minus 6%. This solved the truncation problem and forecasters were able to spread the probability on a range of values that reflected their views as it is clear from Figure (6)[7]. The effect of the truncation is problematic for the analysis in this paper since it clusters probability mass on the lowest open interval and lowers the dispersion of the distribution in quarter 1, which subsequently increases in quarter 2 due to the new grid used in administering the Survey. In the following Section I investigate further this issue in the ECB-SPF as well as in the US-SPF and discuss possible remedies.

4.3 The problem at the open intervals and a solution

To evaluate to what extent the interval boundary constitutes a problem for the real GDP density forecasts in the US-SPF and ECB-SPF, I calculate the percentage of forecasters that assigned a positive probability to the lowest open interval as well as the the average probability that is assigned to the interval[8]. In the first quarter of 2009 a large fraction of forecasters

[4] http://www.ecb.europa.eu/stats/prices/indic/forecast/shared/files/SPF_rounds_dates.pdf?845ed4bf360b41b77415fb1d6a9129f7

[5] http://ec.europa.eu/eurostat/documents/2995521/5067054/2-05032009-AP-EN.PDF/627dcf20-9dd0-4877-874a-c409c68b07d0

[6] The Figure shows the first 20 forecasters out of the 44 that contributed in both quarters sorted by the highest to lowest Bayesian weight calculated as the ratio of prior and posterior precisions.

[7] Only two forecasters assigned probability to the open interval in 2009Q2 with the highest value equal to 12%.

[8] For the US-SPF the lowest open interval for GDP is "< -2" from the first quarter of 1992 until the first quarter of 2009, and was then moved to "< -3" starting from the second quarter of 2009 until present. In the ECB-SPF survey the lowest open interval has been changed several times: "< 0" until the fourth quarter of 2008, then "< -1" until the first quarter of 2009, followed by "< -6" from the second quarter of 2009 until

assigned probability to the lowest open interval in both Surveys. This is particularly the case for the ECB-SPF, since all forecasters assigned positive probability that real GDP for 2009 would be in the interval "< -1" with an average probability of approximately 70%. Instead, for the US-SPF over 80% of the forecasters assigned probability to the lowest interval of "< -2" with an average probability close to 40%. Except for the first quarter of 2009, there are no other quarters in the US-SPF with a a large (average) probability assigned to the left open interval, but there are several instances for the ECB-SPF survey. Since the first quarter of 2012 and continuing into 2013 over half of the forecasters of real GDP for the euro-area assigned probability to the lowest bin, although the average probability is smaller than 10%. This analysis shows that the truncation of the density forecasts is a problem that occurs rarely, but it does happen as it was the case in the first quarter of 2009. The effect of the truncation is to bias the mean and variance obtained from the density forecasts. In addition, the truncation complicates the analysis over time of the characteristics of the individual densities, as it is the case in this paper. To overcome this problem, I propose a simple approach to produce pseudo histograms that represent the density the forecaster would have produced if the interval grid included additional bins beyond the boundary used in the Survey. I will discuss the proposed solution in the specific context of the ECB-SPF histograms reported in the first quarter of 2009, although the solution is more general and applies also to truncation occurring on the right tail of the distribution.

One aspect of the approach that I propose is to use both the point and density forecasts that survey participants produce in the same quarter and for the same target date. The advantage of using the point forecasts is that they are not affected by the boundary problem as it is the case for the density forecasts. In the first quarter of 2009 there were 55 forecasters that contributed points forecasts to the ECB-SPF, with only 5 of them expecting real GDP growth larger than -1%, and the remaining 50 forecasters that expected growth for 2009 to range between -3.2% and -1%[9]. I thus use the point forecasts to determine the center of the pseudo probability distribution. This assumption seems reasonable and it is also supported by the empirical evidence in Engelberg *et al.* (2009) and Clements (2009) that the point forecasts are, to a large extent, consistent with the center of the distribution for most forecasters. Once we have anchored the center of the pseudo distribution, the following step is to characterize its dispersion and shape. I assume that the pseudo distribution follows a triangular distribution. This distribution is simple to handle, but also realistic in this context since it is bounded on a finite interval and allows for asymmetry that, as discussed earlier, is an important feature of survey density forecasts. The triangular distribution is characterized by three parameters: the lower and upper bounds of the support denoted by a and b, and c that locates the mode of the distribution. In the Online Appendix I discuss the details of the construction of the pseudo histograms and discuss the performance of the method.

---

the first quarter of 2010 when the grid went back to "< -1".

[9]Instead, for the US-SPF in that same quarter only half of the forecaster had point forecasts lower than the -2% boundary and with only 10 forecasters (out of 40) assigning a probability between 50 and 75% to the open interval. In the following quarter, only 14 forecasters had point forecasts lower than the new boundary value of -3% while 29 forecasters had point forecasts in the interval between -3% and -2%. The remaining 4 forecasters expected GDP growth to be larger than -2%.

4.4 Weights of the prior

Figure (7) shows the average over time of the Bayesian weight for each forecaster and each quarter for inflation and output growth in the US-SPF and ECB-SPF. The weight is calculated using the pseudo-density forecasts so that these results are not affected by the truncation problem discussed earlier. Values of the weight larger than one indicate that the forecaster was (on average) more uncertain in the current quarter relative to the previous one, despite being one quarter closer to the target date and having observed additional news about the variable and the state of the economy. For both variables and Surveys there are several instances in which the (average) Bayesian weight is larger than one, most often in the second and third quarters, but for some forecasters also in the fourth quarter. The Figure seems to suggest that there are two typical patterns for the evolution of the Bayesian weight. In one case, the average weight is larger than one in the second quarter while later, in most cases, it is below one. On the other hand, several forecasters have a large value of the weight in the third quarter, but not in the second or fourth quarter.

There are two possible explanations for these findings. The first is that forecasters underestimate future uncertainty which they then revise upward in later quarters. The trigger for the non-Bayesian behavior could be the most recent data release that surprises the forecaster, or revisions of earlier released data that causes the forecaster to re-assess the prior density forecast. Another possible explanation is that this is simply the outcome of the binning, a problem which is more likely to occur when the forecaster assigns probability to few bins. This can happen when the individual shifts the location of the forecast which, once discretized to the Survey grid, appears to indicate that uncertainty has increased relative to the previous quarter. An example is a forecaster participating to the US-SPF that in the second quarter assigns 100% probability to the event that GDP growth will be between 2 and 3%. After macroeconomic news is released in the third quarter, the forecaster shifts the location of the density by 0.5% and now believes that there is 100% chance that growth will be between 2.5 and 3.5%. However, the US-SPF requires forecasters to report probabilities in the 2-3% and 3-4% intervals which will require the individual to report a 50% probability to each interval. In addition, the problem of binning a (possibly) continuous distribution should affect differently the two Surveys since the US-SPF has a bin size of 1% whilst the ECB-SPF of 0.5%. To evaluate the effect of the discreteness of the grid, Figure (8) provides a scatter plot of the Bayesian weight in a certain quarter against the number of bins used by the same forecaster for the prediction reported in the previous quarter. It is clear that values of the weight larger than 1 are more likely to occur when the density in the previous quarter is concentrated in a few bins. If the quarters in which a forecaster assigned probability to only one or two bins in the previous quarter are excluded, the average Bayesian weight shows still many cases of weights larger than 1 in both Surveys. This exploratory analysis thus suggests that discreteness of the grid and the bin size adopted in the US-SPF and ECB-SPF might play a role, but not fully explain the non-Bayesian updating by some forecasters.

Overall, the analysis so far has indicated that, at times, individuals update their density forecasts by expecting more uncertainty, despite being closer to the target date and having observed additional macroeconomic news about the state of the economy. There are several possible explanations for this behavior. One is that forecasters might have weak incentives to report accurate density forecasts, as opposed to their point forecasts that receive significant scrutiny from the public. Stark (2013) conducted a Survey of US-SPF participants and found that only 8 forecasters use the density forecasts in their analysis while 17 forecasters produce the forecasts only for the US-SPF. In addition, 11 participants declared to use the results of the survey's density forecasts in their work as opposed to 15 who do not. This suggests that the finding that, occasionally, weights are larger than 1 might be the outcome of inattention when it comes to predict the dispersion of the forecast distribution. If this is the case, then I would expect that the weights are not affected by macroeconomic news in any systematic way. Another possible explanation for this result is related to the discretization of the survey grid that might lead to spurious findings of increased uncertainty in the revised density forecasts. In this case, the occurrence of non-Bayesian updating should be related to the occurrence of density forecasts concentrated on a few bins, with fewer of them increasing the probability of the event. The third explanation is motivated by the theoretical models of Epstein (2006), Ortoleva (2012), and Nimark (2014) that interpret the inconsistent updates as the result of news that trigger a re-evaluation of the beliefs about uncertainty.

In order to investigate the relevance of these explanations, in the following Section I investigate the factors that drive the variation over time of the Bayesian weight and the probability of non-Bayesian behavior. To this goal, I construct a measure of individual surprise in the macroeconomic announcement about the variable, relative to the forecaster's prior quarter belief. In addition, I include in the analysis the number of bins that were used by forecasters in the prediction of the previous quarter to evaluate the effect of discreteness on causing non-Bayesian behavior.

5 Empirical evidence

The aims of this Section are 1) to test the inequality predicted by the BLM that the precision should not decrease as the quarters advance, and (2) to find the factors that drive the variation of the Bayesian weight. In particular, I will consider the unexpected news and the number of bins used in the previous quarter as the potential factors driving the weights over time and across forecasters. The surprise measure should be large when the released data is in the tails of the prior distribution, and small when the data released is close to the center of the forecast distribution.

5.1 Testing the inequality across horizons

Patton and Timmermann (2012) propose to test for forecast rationality by exploiting the term structure of forecasts, rather than focusing on a single horizon. They derive alternative implications of forecast rationality, such as that the mean square forecast error should be increasing with the horizon. The methodology of Patton and Timmermann (2012) can be used to test the prediction of the BLM that the precision of the density forecast should be declining with the horizon and increasing with the quarter, q, that is $\psi_{q-1,t} \leq \psi_{q,t}$ (for $q = 2, 3, 4$).

To test these restrictions I implement the univariate optimal revision regression approach that consists of a regression of the short-horizon forecast on the long-horizon forecast and the forecast revisions, that is,

$$\psi_{4,t} = \alpha + \beta_1 \psi_{1,t} + \sum_{q=2}^{3} \beta_q \phi_{q,t} + \eta_t$$

where $\psi_{4,t}$ and $\psi_{1,t}$ are the forecast precisions in quarter 4 and 1, respectively, and $\phi_{q,t} = \psi_{q,t} - \psi_{q-1,t}$ represents the precision revision and, in the context of the BLM, the precision of the signal as in Equation (3) Notice that in the model above the dependent variable is the expected precision rather than its realization which is unobserved. This regression represents a Mincer-Zarnowitz (MZ) type of regression where the goal is to test the hypothesis:

$$H_0 : \alpha = 0 \cap \beta_q = 1 \text{ for q = 1,2,3}$$

Given the panel structure of the dataset, I implement the test in two ways: one by pooling all forecasters and the other by running the test on each forecasters. Table (2) reports the p-value of the test for the pooled case and the fraction of rejection of the null hypothesis using a 5 and 10% significance level across forecasters. The results indicate strong rejections for both variables and surveys which confirm, statistically, the evidence of departures from the BLM prediction of increasing precision as the horizon declines. When we apply the test to the individual forecasters the results show rejections for between 50 to 80% of the forecasters. Overall, these results indicate that forecasters do not update their density forecast in a Bayesian manner, in particular in the sense of becoming less uncertain about the variable closer to the end of the year and more information is available.

5.2 Empirical specification

The empirical specification aims at explaining the Bayesian weight using a measure of surprise relative to the subjective density of a forecaster[10] and the number of bins in the previous quarter density forecast. The surprise measure can be interpreted as a standardized measure of the data release in the current quarter relative to the mean and standard deviation of the density forecast in the previous quarter. A large (absolute) value of the surprise indicates that the newly released data was considered unlikely by the forecaster based on the previous quarter density forecast. The surprise measure is similar to the uncertainty index proposed by Rossi and Sekhposyan (2015) which aims at measuring how extreme a realization is relative to a certain distribution. I consider the possibility that the Bayesian weight of forecaster i in quarter q of year t, denoted $\rho_{i,q,t}$, might react differently to positive or negative surprises. I achieve this by creating the variables $S^+_{i,q,t}$ and $S^-_{i,q,t}$ that are equal to the individual surprise $S_{i,q,t}$ when the surprise is positive and negative, respectively, and zero otherwise. The second variable that I include in the model is $Bins_{i,q-1,t}$ that represents the number of bins that

[10]The details of the construction of the surprise measure are available in the Online Appendix.

forecaster i assigned probability to in the previous quarter. The reason for including this variable is to control for the possibility that the discreteness of the histogram might be the reason of the non-Bayesian behavior. In addition, in all regression models I include quarter fixed effects to account for systematic differences in the level of the weight across forecasters. I denote the dummy variable for the third and fourth quarter in quarter q of year t by $Q3_{q,t}$ and $Q4_{q,t}$, respectively. I also control for the possibility of persistence or reversal in the dynamics of the Bayesian weight by including its value in the previous quarter, $\rho_{i,q-1,t}$.

I consider two specifications of the panel model, one in which the effect of the explanatory variables are common across all forecasters and the other in which the effects are homogeneous to a group of forecasters but heterogeneous across groups. In both cases, the quarterly dummy effects are forecaster-specific. Denoting by $X_{i,q,t}$ the vector of explanatory variables for forecaster i in quarter q of year t, the pooled FE model is defined as

$$\rho_{i,q,t} = \beta_{i,3}Q3_{q,t} + \beta_{i,4}Q4_{q,t} + X^{'}_{i,q,t}\gamma + \epsilon_{i,q,t} \tag{8}$$

where the βs are forecaster-specific while the coefficient γ is common across all forecasters. I also consider a model in which the parameters are allowed to be heterogenous across different groups of forecasters. The grouped FE model has recently been proposed by Lin and Ng (2012) as a way to account for heterogeneity among individuals in a parsimonious way. Assume that there are N forecasters and each belong to one of G groups. The grouped FE model for the Bayesian weight is given by:

$$\rho_{i,q,t} = \beta_{i,3}Q3_{q,t} + \beta_{i,4}Q4_{q,t} + X^{'}_{i,q,t}\gamma_g + \epsilon_{i,q,t} \tag{9}$$

for $i \in g$ and $g = 1, \cdots, G$. The effect of the independent variables γ_g are common across forecasters that belong to the same group, but different across groups. I estimate the model using the K-means algorithm of Lin and Ng (2012) and select G, the optimal number of groups, using the modified BIC criterion with $\sqrt{min(N, \overline{T})}$ penalty, where N is the number of forecasters and $\overline{T}$ is the average number of time periods. This penalty makes the criterion asymptotically consistent in the presence of estimation uncertainty.

I start by estimating the models by OLS and test the null hypothesis of no cross-sectional dependence. In case of rejection of the null hypothesis, I re-estimate the model using the Common Correlated Estimator (CCE) proposed by Pesaran (2006) that provides consistent estimates in the presence of unknown common shocks. I test for the presence of cross-sectional correlation in the errors using the Cross-Section Dependence (CD) test proposed by Pesaran (2004). The test statistic is standard normally distributed and is given by

$$CD = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{i,j}}\hat{\xi}_{i,j} \right) \tag{10}$$

where $\xi_{i,j}$ represents the correlation coefficient of the errors in Equations (8) or (9) for forecaster

i and j, $T_{i,j}$ represents the number of observations available for each pair of forecasters (which might differ due to the unbalanced nature of the panel), and N is the number of forecasters. In the application, I found that, for both variables and Surveys, the null hypothesis of cross-sectional independence is rejected for the OLS residuals so that all estimation results reported refer to the CCE. In addition, I test the homogeneity of the slope parameters across individuals, both in the pooled FE model, and within each group in the grouped FE model. I use the $\tilde{\Delta}$ dispersion test proposed by Pesaran and Yamagata (2008) defined as $(\tilde{S} - k\sqrt{N})/\sqrt{2k}$ that is standard normal distributed, with k denoting the number of parameters being tested and N the number of forecasters. The quantity $\tilde{S}$ is provided in Equation (13) of Pesaran and Yamagata (2008) and represents a weighted average of the square distance of the individual parameter estimates from the weighted FE pooled estimator.

5.3 Estimation results

Tables (3)-(4) and (6)-(5) provide the estimation results for the pooled and grouped FE model for the US-SPF and the ECB-SPF, respectively. The results of the homogeneity test indicate that, with the exception of US-SPF GDP, the null of parameter homogeneity is rejected in the pooled FE model, but it is not rejected for the individual groups. This indicates that the group FE model is able to distinguish the different characteristics of the forecasters and cluster them in groups that are homogeneous in the reaction of the Bayesian weight to surprises, number of bins, and the lagged value variable. The BIC criterion indicates an optimal choice of G between 2 and 3 groups for all variables, although adding more groups seems to produce large increases in the goodness-of-fit statistic and statistically significant estimates. This can be explained by the heavy penalization used by the BIC criterion that discourages the inclusion of more groups and makes the BIC criterion of the group FE model being slightly larger relative to the pooled FE model in some cases. However, I do not consider this to be worrisome since the differences are extremely small and the coefficient estimates indicate significant heterogeneity among forecasters. In terms of the cross-sectional dependence in the residuals, there are a few cases where the CD statistic rejects the null hypothesis of independence of the residuals even though the models are estimated with the CCE. However, the magnitudes of the average correlation of the residuals are between 0.02 and 0.07 in absolute value that are quite small to create significant bias in our estimates.

The results for the US-SPF PGDP in Table (3) indicate that all variables are significant in the pooled case and the presence of 3 groups. The largest is a group of 15 forecasters that are not sensitive to news, although they show a significant negative relationship between the Bayesian weight and the number of bins in the previous quarter. This suggests that this group of forecasters increases the dispersion of their density forecasts following quarters in which they produced narrow forecasts. This is consistent with the earlier evidence that, at least in part, the non-Bayesian behavior might be due to a discretization effect in which forecasters adjust their density forecast following quarters in which they produced narrow density forecasts. I also find that there is a tendency for this group to revert the weight relative to the previous period. The remaining 7 forecasters are split between a group of two forecasters that respond only to news, and a group of 5 forecasters that react to all the factors included in the model.

The opposite sign of the estimated coefficient for positive and negative surprises indicates that both good and bad news increase the Bayesian weight and the dispersion of the density forecasts. For this group of forecasters the non-Bayesian behavior seems to be explained by the effect of large surprises that causes them to re-evaluate their density forecasts and decrease the precision of their forecasts.

I find that also for the US-SPF GDP (in Table 4) 3 groups represent a good compromise between goodness-of-fit and the number of parameters in the model. In this case, a large group of 18 forecasters react to positive news about output and to the number of bins. Positive surprises about GDP contribute to increase the Bayesian weight for a small group of 3 forecasters, while forecaster 65 forms a group of its own. Although the estimates for this forecaster are not statistically significant, the large coefficients for the news variables indicate that the forecaster was unique in its reaction to news, relative to the other forecasters. This also shows an advantage of the group estimator that is able to group forecasters, but also not to group them when their behavior is substantially different from the other individuals. Overall, the evidence for GDP in the US-SPF confirms the results for PGDP that both the reaction to surprises and to narrow density forecasts in earlier quarters explain large values of the Bayesian weight.

In the case of the ECB-SPF Survey, I find that two groups are optimal for GDP and three groups for HICP. For both these variables, the largest group (19 for GDP and 27 for HICP) react to changes in all variables with the direction of the relationship similar to the earlier findings for US-SPF. For GDP, a group of 12 forecasters seems to react to negative surprises, but not positive. It could be the case that these forecasters increase the dispersion of their forecast after they miss the quarterly forecast by being too optimistic. Instead, for HICP one of the two smaller groups is composed of 3 forecasters that react to all variables, although the magnitude of the effects are significantly different from the large group. The third group includes only forecaster 52 which reacts to both positive and negative surprises, but seems not to change its weight based on the number of bins used in the previous quarter.

6 Conclusion

Are professional forecasters Bayesian? The findings in this paper show that occasionally forecasters update their density forecasts in a way that is inconsistent with Bayesian learning. The forecasts that I consider in this paper are for a fixed target date (i.e., GDP growth in a certain year) so that the release of macroeconomic news should lead forecasters to expect a decline in uncertainty as the forecast date approaches the target date. I find that this prediction is sometimes violated by most individuals since they provide density forecasts that are more disperse relative to the one they provided in the previous quarter. I identify three possible explanations for this finding.

The first is a methodological issue with the way the Surveys are conducted. Both the US-SPF and ECB-SPF ask forecasters to assign probabilities that the variables of interest will fall in a predetermined set of bins, with the first and last being open intervals. Although the grid is set wide enough to include the distribution of most forecasters, it is possible that unexpected economic events might rapidly change the views of forecasters to the point that they might assign a large probability to the open intervals. This was the case in the first quarter of 2009 for the ECB-SPF. Forecasters became very pessimistic about the growth outlook of the European economy and expected a severe decline in output. However, the grid used in the Survey in that quarter had the lowest open bin set at -1% or lower which received most of the probability for all forecasters. In the following quarter the grid was extended to -6% and lower which solved the truncation problem. However, when comparing the truncated distributions of 2009Q1 to the forecasts of 2009Q2 it appears as if forecasters were more uncertain in the second quarter simply because of the spuriously low uncertainty extracted from the truncated distributions. To solve this problem, I propose an approach to construct pseudo-density forecasts that exploits the fact that each forecaster provides both a density and point forecasts, with the latter being unaffected by the boundary issue.

Once I resolve the boundary problem, I still find several instances in which forecasters in the US-SPF and ECB-SPF increase the dispersion of their density forecasts relative to their previous quarter forecasts. I find two main channels that seem to explain, at least partly, this behavior. The first mechanism arises when forecasters are surprised by the data release. The empirical evidence indicates that a large value (either positive or negative) is likely to lead the forecaster to re-evaluate the evidence and, in some cases, to revise the prior density that the forecaster is updating. Another situation in which professional forecasters are likely to update their prior in a non-Bayesian manner is when they produce narrow forecasts in the early quarters of the target year. Also in this situation, forecasters tend to produce posterior distributions that are more disperse relative to the prior distribution. This suggests that also the discretization of a continuous distribution into bins might explain some of the findings. Another result of the analysis is that forecasters seem to cluster in different groups that react differently to surprises and the number of bins that they used in the previous quarter. I find significant heterogeneity across forecasters, with some groups insensitive to surprises while others reacting only to positive and/or negative unexpected data releases. Similarly, some groups of forecasters seem to revise their density forecasts in response to having produced a narrow forecasts, while others do not.

The answer to the question above is thus that professional forecasters update their densities in a Bayesian manner, although they sometimes deviate from this behavior, mostly because they underestimate uncertainty at long forecast horizons. As forecasters get closer to the target date, they become more attentive to real-time news and evaluate more accurately their density forecasts. At times, this leads forecasters to expect higher uncertainty about the realization of the macroeconomic variable. Overall, the analysis indicates that forecasting economic variables is a difficult task that becomes even more challenging when the goal is to forecast the whole distribution of possible outcomes.

References

Abel, Joshua, Rich, Robert, Song, Joseph and Tracy, Joseph (2016). The measurement and behavior of uncertainty: Evidence from the ECB Survey of Professional Forecasters. *Journal of Applied Econometrics*, **31,** 533–550.

Andrade, Philippe and Le Bihan, Hervé (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, **60,** 967–982.

Boero, Gianna, Smith, Jeremy and Wallis, Kenneth F (2015). The measurement and characteristics of professional forecasters' uncertainty. *Journal of Applied Econometrics*, **30,** 1029–1046.

Bowles, Carlos, Friz, Roberta, Genre, Veronique, Kenny, Geoff, Meyler, Aidan and Rautanen, Tuomas (2007). The ECB Survey of Professional Forecasters (SPF)- A review after eight years' experience. *ECB Occasional Paper*, number 59.

Brock, William A and Hommes, Cars H (1997). A rational route to randomness. *Econometrica*, **65,** 1059–1095.

Brock, William A and Hommes, Cars H (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, **22,** 1235–1274.

Clements, Michael P (2009). Internal consistency of survey respondents? forecasts: Evidence based on the survey of professional forecasters. *The methodology and practice of econometrics. A festschrift in honour of David F. Hendry*, 206–226.

Clements, Michael P (2010). Explanations of the inconsistencies in survey respondents' forecasts. *European Economic Review*, **54,** 536–549.

Clements, Michael P (2014). Forecast uncertainty ex ante and ex post: US inflation and output growth. *Journal of Business & Economic Statistics*, **32,** 206–216.

Coibion, Olivier and Gorodnichenko, Yuriy (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, **105,** 2644–2678.

Croushore, Dean, Stark, Tom *et al.* (2019). Fifty years of the survey of professional forecasters. *Economic Insights*, **4,** 1–11.

D'Amico, Stefania and Orphanides, Athanasios (2008). Uncertainty and disagreement in economic forecasting. *FRB working paper.*

De Bruin, Wändi Bruine, Manski, Charles F., Topa, Giorgio and van der Klaauw, Wilbert (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics*, **26,** 454–478.

Del Negro, Marco, Casarin, Roberto and Bassetti, Federico (2018). A bayesian approach for inference on probabilistic surveys. Technical Report. Working Paper.

Dovern, Jonas, Fritsche, Ulrich and Slacalek, Jiri (2012). Disagreement among forecasters in G7 countries. *Review of Economics and Statistics*, **94,** 1081–1096.

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

Epstein, Larry G (2006). An axiomatic model of non-Bayesian updating. *The Review of Economic Studies*, **73,** 413–436.

Epstein, Larry G, Noor, Jawwad and Sandroni, Alvaro (2008). Non-Bayesian updating: a theoretical framework. *Theoretical Economics*, **3,** 193–229.

Epstein, Larry G, Noor, Jawwad and Sandroni, Alvaro (2010). Non-Bayesian learning. *The BE Journal of Theoretical Economics*, **10**.

Ganics, Gergely, Rossi, Barbara and Sekhposyan, Tatevik (2020). From fixed-event to fixed-horizon density forecasts: Obtaining measures of multi-horizon uncertainty from survey density forecasts.

Garcia, Juan A (2003). An introduction to the ECB's Survey of Professional Forecasters. *ECB Occasional Paper*, number 8.

Garcia, Juan A and Manzanares, Andres (2007). What can probability forecasts tell us about inflation risks? *ECB Occasional Paper*, number 825.

Giordani, Paolo and Söderlind, Paul (2003). Inflation forecast uncertainty. *European Economic Review*, **47,** 1037–1059.

Kandel, Eugene and Pearson, Neil D (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, **103,** 831–72.

Kandel, Eugene and Zilberfarb, Ben-Zion (1999). Differential interpretation of information in inflation forecasts. *Review of Economics and Statistics*, **81,** 217–226.

Kendall, Maurice and Stuart, Alan (1977). The advanced theory of statistics, vol. 1. *New York: MacMillan*.

Kenny, Geoff, Kostka, Thomas and Masera, Federico (2014). How informative are the subjective density forecasts of macroeconomists? *Journal of Forecasting*, **33,** 163–185.

Kenny, Geoff, Kostka, Thomas and Masera, Federico (2015). Density characteristics and density forecast performance: a panel analysis. *Empirical Economics*, **48,** 1203–1231.

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

Lahiri, Kajal and Sheng, Xuguang (2008). Evolution of forecast disagreement in a Bayesian Learning Model. *Journal of Econometrics*, **144,** 325–340.

Lahiri, Kajal and Wang, Wuwei (2020). Estimating macroeconomic uncertainty and discord using info-metrics. In *Innovations in Info-metrics: A Crossdisciplinary perspective on Information and Information processing* (eds Min Chen, Michael Dunn, Amos Golan and Aman Ullah). Oxford University Press.

Lin, Chang-Ching and Ng, Serena (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, **1,** 42–55.

Mankiw, N Gregory and Reis, Ricardo (2002). Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. *Quarterly Journal of Economics*, **117,** 1295–1328.

Mankiw, N Gregory, Reis, Ricardo and Wolfers, Justin (2004). Disagreement about inflation expectations. In *NBER Macroeconomics Annual 2003, Volume 18*, pp. 209–270. The MIT Press.

Manzan, Sebastiano (2011). Differential interpretation in the Survey of Professional Forecasters. *Journal of Money, Credit, and Banking*, **43,** 993–1017.

Nimark, Kristoffer (2014). Man-bites-dog business cycles. *American Economic Review*, **104,** 2320–2367.

Ortoleva, Pietro (2012). Modeling the change of paradigm: Non-Bayesian reactions to unexpected news. *The American Economic Review*, **102,** 2410–2436.

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

Patton, Andrew J and Timmermann, Allan (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, **30,** number 1, 1–17.

Pesaran, M. Hashem (2004). General diagnostic tests for cross section dependence in panels. *Working paper.*

Pesaran, M. Hashem (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, **74,** 967–1012.

Pesaran, M Hashem and Yamagata, Takashi (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, **142,** 50–93.

Rich, Robert and Tracy, Joseph (2010). The relationships among expected inflation, disagreement, and uncertainty: evidence from matched point and density forecasts. *The Review of Economics and Statistics*, **92,** 200–207.

Rossi, Barbara and Sekhposyan, Tatevik (2015). Macroeconomic uncertainty indices based on nowcast and forecast error distributions. *American Economic Review*, **105,** number 5, 650–55.

Sheng, Xuguang and Yang, Jingyun (2013). Truncated product methods for panel unit root tests. *Oxford Bulletin of Economics and Statistics*, **75,** 624–636.

Sims, Christopher A (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50,** 665–690.

Stark, Tom (2013). SPF panelists' forecasting methods: A note on the aggregate results of a November 2009 Special Survey. *working paper.*

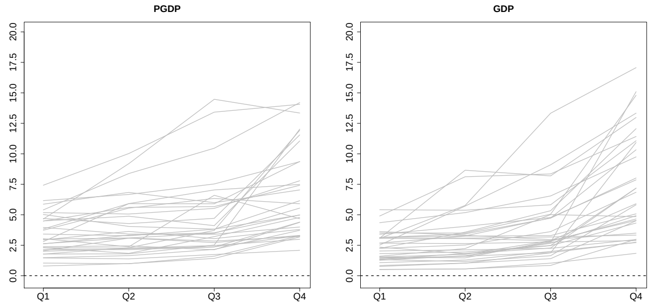
Tvesky, Amos and Kanheman, Daniel (1974). Judgement under uncertainty: heuristics and biases. *Science*, **185,** 1124–1131.

Woodford, Michael (2002). Imperfect common knowledge and the effects of monetary policy. In *Knowledge, Information and Expectations in Modern Macroeconomics* (eds P. Aghion, R. Frydman, J. Stiglitz and M. Woodford). Princeton University Press.

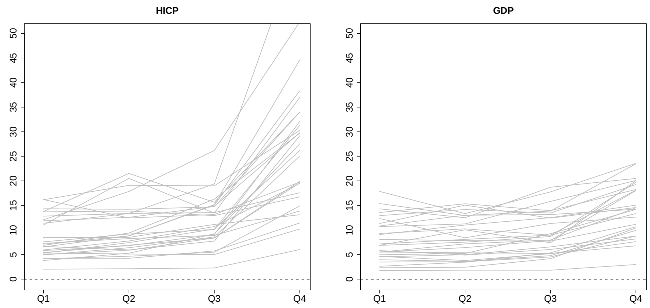
Zarnowitz, Victor and Lambros, Louis A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, **95,** 591–621.

Figure 1: **Average Precision by Forecaster and Quarter**

US-SPF



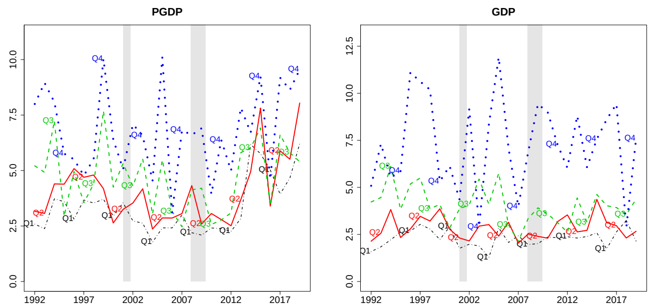
ECB-SPF



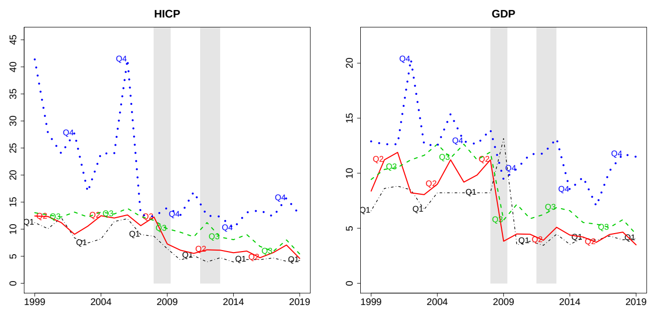
Each line represents the time average of the precision (the inverse of the variance) of a forecaster's density forecasts by quarter. The top two graphs refer to the US-SPF and the bottom graphs refer to the ECB-SPF.

Figure 2: **Average Precision by Quarter over Time**

US-SPF



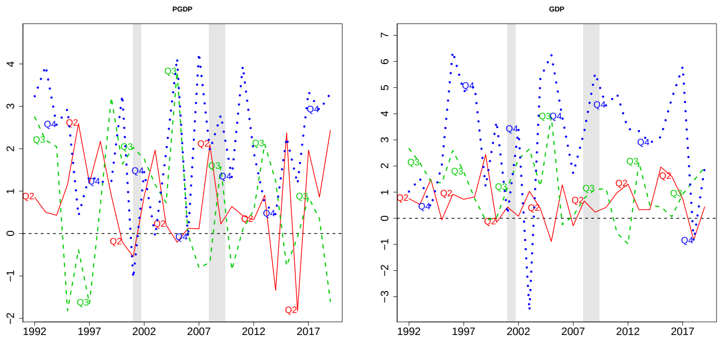
ECB-SPF



Cross-sectional average of the precision of the density forecasts in each quarter. The shaded areas represent the NBER and CEPR recession periods. The top two graphs refer to the US-SPF and the bottom graphs refer to the ECB-SPF.

Figure 3: **Signal Precision by Quarter over Time**

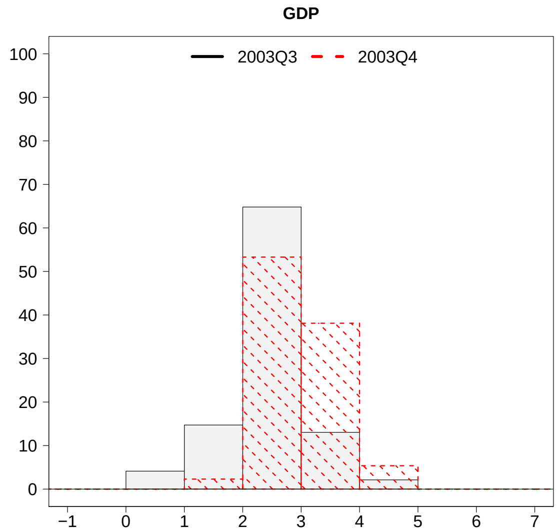
US-SPF



ECB-SPF

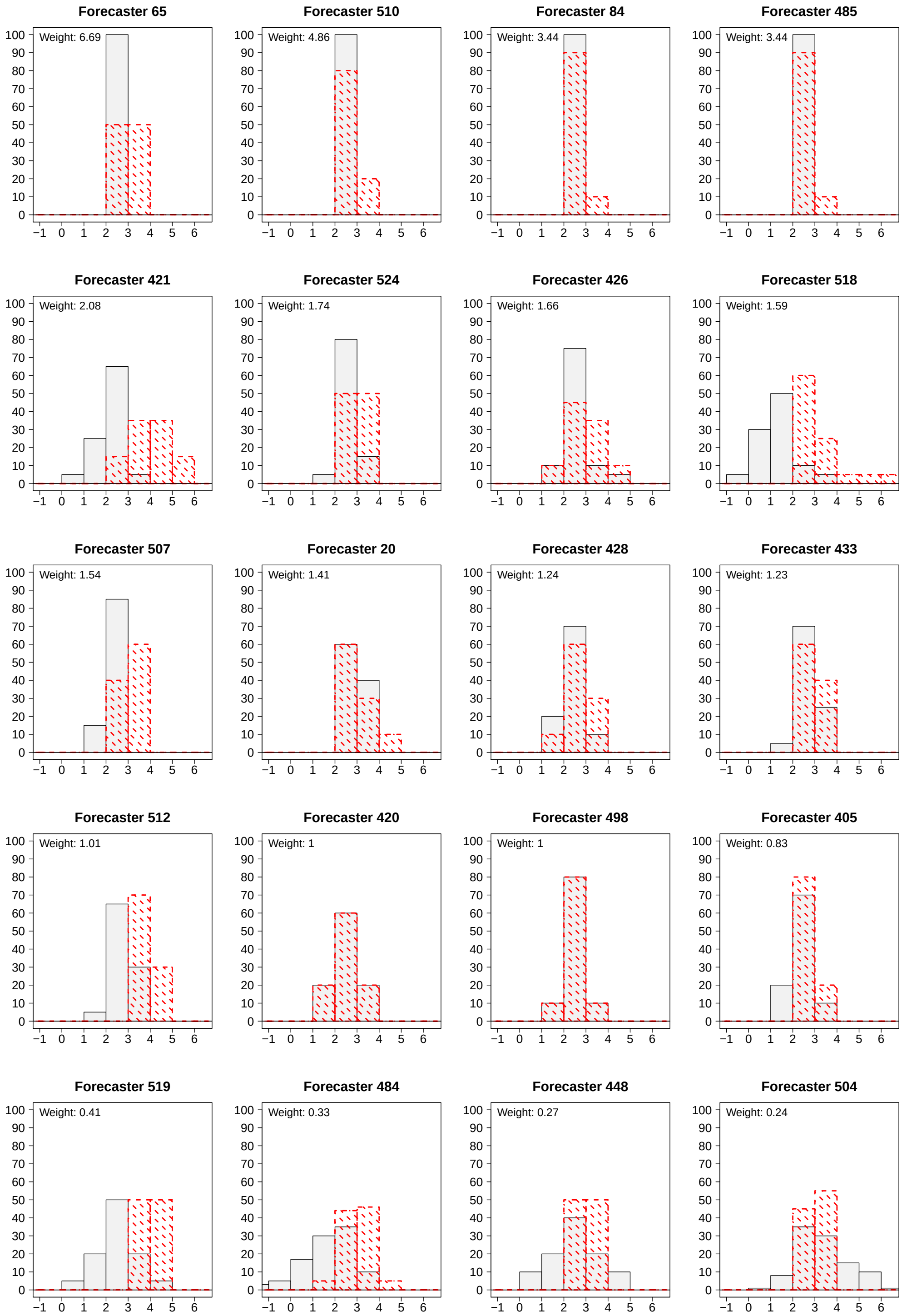
Cross-sectional average of the signal precision obtained as the difference between the prior and posterior precisions in each quarter. The shaded areas represent the NBER and CEPR recession periods. The top two graphs refer to the US-SPF and the bottom graphs refer to the ECB-SPF.

Figure 4: **US-SPF Consensus Forecasts in 2003Q3 and 2003Q4**



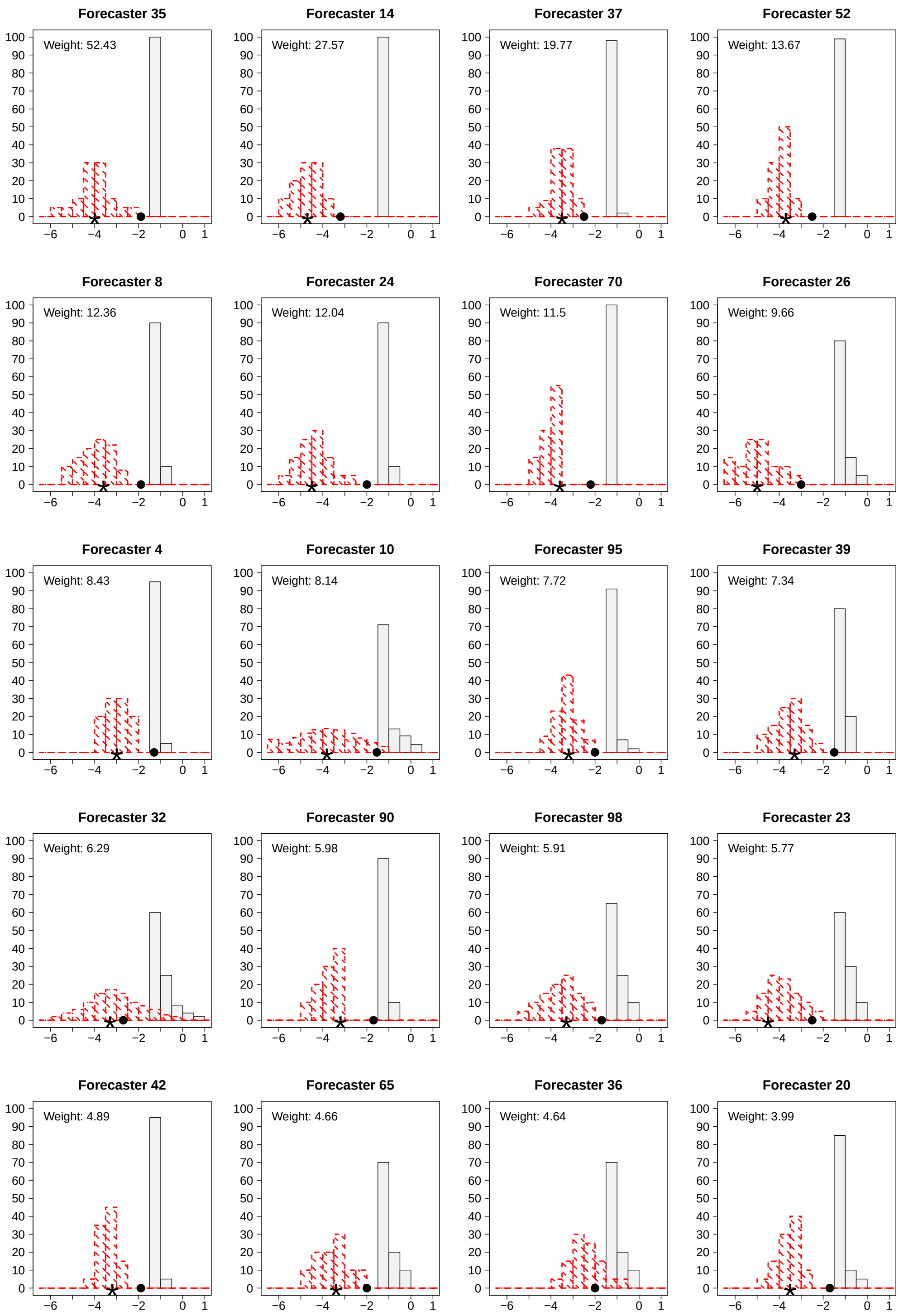
Consensus density forecasts for real GDP growth in the third and fourth quarter of 2003. The consensus forecast is obtained by averaging the probabilities reported in each bin by 25 forecasters in quarter 3 and 30 in quarter 4.

Figure 5: **US-SPF Forecasters in 2003Q3 and 2003Q4**



Individual density forecasts for 20 forecasters that provided density forecasts in quarter 3 and 4 of 2003. The shaded bars represent 2003Q3 and the bar with red dashed lines are for 2003Q4. The weight provided in the top-left corner of the graph represents the ratio of the prior and posterior precisions. A value of the weight larger than 1 indicates the forecaster has increased the uncertainty of the distribution in quarter 4 relative to the previous quarter.

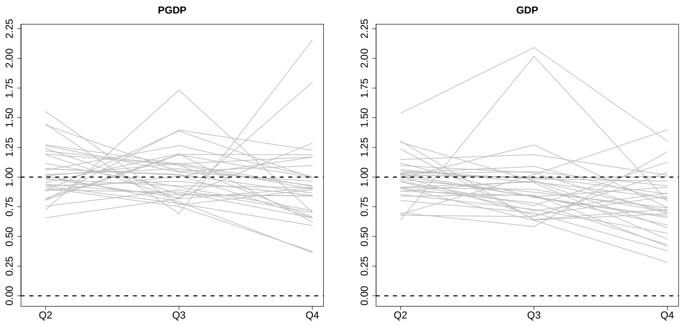
Figure 6: **ECB-SPF Forecasters in 2009Q1 and 2009Q2**



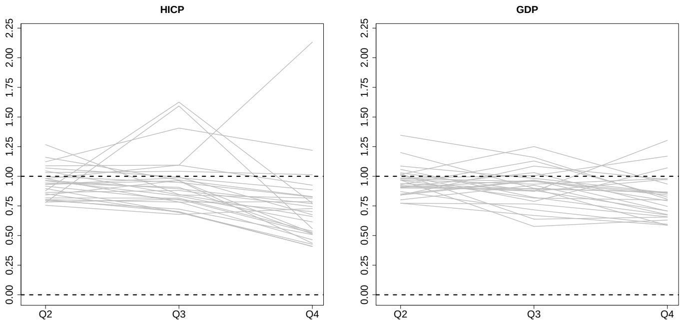
Individual density forecasts for 20 forecasters that participated to the Survey in quarter 1 and 2 of 2009. The histograms for 2009Q1 is indicated by the gray bar and the point forecast of each forecaster is denoted by a dot in the x-axis, while the histogram for 2009Q2 is denoted by bar with red dashed lines and the point forecasts are denoted by a star. The weight defined as the ratio of the prior and posterior precisions are provided in the top left corner of the graph. A value of the ratio larger than 1 indicates the forecaster increased the uncertainty of the distribution in quarter 2 relative to quarter 1 of 2009.

Figure 7: **Average Bayesian Weight by Forecaster and Quarter**

US-SPF



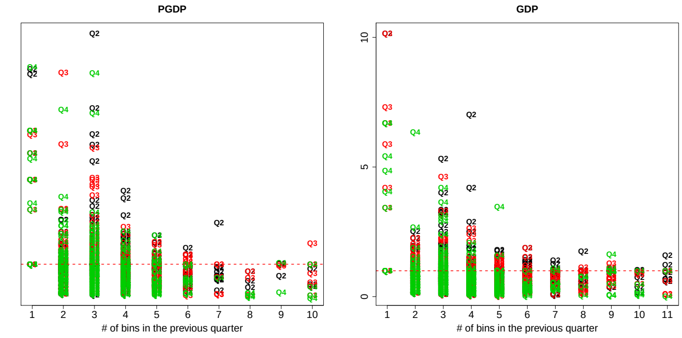
ECB-SPF



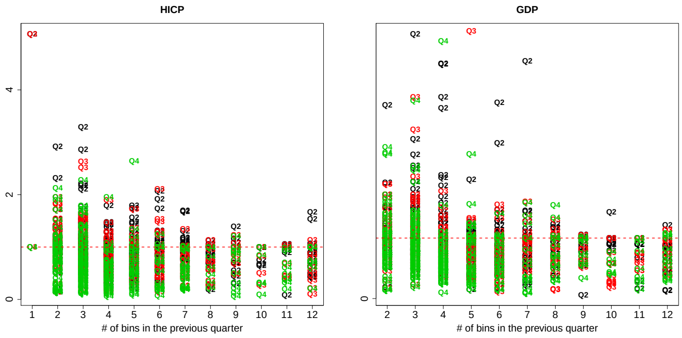
Each line represents the time average of a forecaster's Bayesian weight by quarter. The weight is defined as the ratio of the posterior to the prior precisions. The precision is calculated on the pseudo-density forecasts that account for the boundary problem. The left graph refers to forecast of the PGDP growth rate for the current year and includes 23 forecasters and the right graph refers to the GDP growth rate and includes 24 forecasters.

Figure 8: **Bayesian Weight** *vs* **# of Bins**

US-SPF



ECB-SPF



Scatter plot of the number of bins used by a forecaster to provide a density forecast in a certain quarter and the Bayesian weight in the following quarter. Labels indicate the quarter of the weight and the green color is used for Q4, red for Q3, and in black for Q2. These graphs refer to the pseudo-denstiy forecasts.

Table 2: **PT Rationality Test**

	US		ECB	
	GDP	PGDP	GDP	HICP
Pooled	0.00	0.00	0.00	0.00
Indiv (5%)	0.45	0.62	0.74	0.77
(10%)	0.51	0.66	0.78	0.80

Results of the Patton and Timmermann (2012) regression test of the inequality $\phi_{q-1,t} \leq \phi_{q,t}$. The values reported are the p-values for the pooled case and the percentage of rejection of the test at the 5 and 10% significance level applied on the individual forecasts.

Table 3: **US-SPF PGDP**

Variable	Pooled	Group 1	Group 2	Group 3
$\rho_{i,q-1,t}$	**-0.15**	**-0.33**	0.26	**-0.14**
p-value	0.00	0.00	0.20	0.00
$S^{+}_{i,q,t}$	**0.97**	**2.89**	**2.49**	0.11
p-value	0.00	0.04	0.00	0.48
$S^{-}_{i,q,t}$	**-0.38**	**-1.38**	**-3.44**	**-0.19**
p-value	0.00	0.00	0.00	0.02
$Bins_{i,q-1,t}$	**-0.09**	-0.05	0.02	**-0.09**
p-value	0.00	0.60	0.94	0.00
N	32	4	2	26
CD	-1.81	-1.66		
Av. Corr.	-0.03	-0.02		
BIC	6.09	6.12		
R-square	0.70	0.77	0.69	0.84
Hom.	4.36	-0.11	-0.69	-2.11
Group 1:	407 456 508 535			
Group 2:	65 84			
Group 3:	20 411 420 421 422 426 428 431 433 446 463			
	72 483 484 504 507 510 512 518 520 524 546 548 549 555 557			

Estimation results for the pooled FE models in Equation (8) (first column) and the grouped FE model in Equation (9) (second to fourth columns). N represents the number of forecasters, CD the cross-sectional dependence test of Pesaran (2004), $Av.\ Corr.$ represents the average cross-sectional correlation of the model residuals, BIC is the modified BIC criterion proposed by Lin and Ng (2012), $R-square$ is the goodness-of-fit statistic, and $Hom.$ represents the $\tilde{\Delta}$ dispersion test for parameter homogeneity proposed by Pesaran and Yamagata (2008). The CD and $Hom.$ tests are standard normal distributed. The bottom of the Table provides the group membership based on the identifier provided in the Survey.

Table 4: **US-SPF GDP**

Variable	Pooled	Group 1	Group 2	Group 3
$\rho_{i,q-1,t}$	**-0.15**	**-0.16**	0.01	-0.22
p-value	0.00	0.00	0.95	0.17
$S^{+}_{i,q,t}$	**0.91**	**0.45**	**1.80**	0.75
p-value	0.00	0.00	0.01	0.09
$S^{-}_{i,q,t}$	-0.18	-0.11	-0.37	0.04
p-value	0.26	0.37	0.79	0.96
$Bin_{i,q,t}$	**-0.07**	**-0.04**	**-0.79**	**-0.55**
p-value	0.00	0.00	0.04	0.01
N	33	29	2	2
CD	0.14	-1.17		
Av. Corr.	0.00	-0.02		
BIC	6.21	6.08		
R-square	0.64	0.76	0.75	0.75
Hom.	2.91	2.12	-1.58	-0.21
Group 1:	20 407 411 420 421 422 426 428 431 433 446 456			
	463 472 483 484 504 507 508 512 518 520 524 535 546 548 549 555 557			
Group 2:	65 527			
Group 3:	84 510			

Estimation results for the pooled FE models in Equation (8) (first column) and the grouped FE model in Equation (9) (second to fourth columns). N represents the number of forecasters, CD the cross-sectional dependence test of Pesaran (2004), $Av.\ Corr.$ represents the average cross-sectional correlation of the model residuals, BIC is the modified BIC criterion proposed by Lin and Ng (2012), $R-square$ is the goodness-of-fit statistic, and $Hom.$ represents the $\tilde{\Delta}$ dispersion test for parameter homogeneity proposed by Pesaran and Yamagata (2008). The CD and $Hom.$ tests are standard normal distributed. The bottom of the Table provides the group membership based on the identifier provided in the Survey.

Table 5: **ECB-SPF HICP**

Variable	Pooled	Group 1	Group 2	Group 3
$\rho_{i,q-1,t}$	**-0.08**	-0.02	0.68	**-0.16**
p-value	0.01	0.39	0.21	0.08
$S^{+}_{i,q,t}$	**0.39**	**0.15**	**1.28**	**1.05**
p-value	0.00	0.00	0.00	0.00
$S^{-}_{i,q,t}$	**-0.57**	**-0.21**	**-7.33**	**-2.06**
p-value	0.00	0.00	0.01	0.00
$Bins_{i,q-1,t}$	**-0.03**	**-0.03**	-0.19	**-0.34**
p-value	0.00	0.00	0.26	0.00
N	35	31	1	3
CD	-0.27	-0.41		
Av. Corr.	0.00	0.00		
BIC	6.38	6.22		
R-square	0.71	0.72	0.84	0.73
Hom.	5.12	1.87	0.00	0.41
Group 1:	4 14 15 16 22 23 24 26 29 31 32 33 37 38 39 41 47 54 56 73 80 84 85 89 90 92 93 94 95 96 98			
Group 2:	52			
Group 3:	20 35 70			

Estimation results for the pooled FE models in Equation (8) (first column) and the grouped FE model in Equation (9) (second to fourth columns). N represents the number of forecasters, CD the cross-sectional dependence test of Pesaran (2004), *Av. Corr.* represents the average cross-sectional correlation of the model residuals, BIC is the modified BIC criterion proposed by Lin and Ng (2012), $R-square$ is the goodness-of-fit statistic, and *Hom.* represents the $\tilde{\Delta}$ dispersion test for parameter homogeneity proposed by Pesaran and Yamagata (2008). The CD and *Hom.* tests are standard normal distributed. The bottom of the Table provides the group membership based on the identifier provided in the Survey.

Table 6: **ECB-SPF GDP**

Variable	Pooled	Group 1	Group 2
$\rho_{i,q-1,t}$	**-0.10**	**-0.19**	**-0.07**
p-value	0.00	0.00	0.04
$S^{+}_{i,q,t}$	**0.32**	**0.66**	**0.16**
p-value	0.00	0.00	0.03
$S^{-}_{i,q,t}$	**-0.42**	**-0.30**	**-0.68**
p-value	0.00	0.00	0.00
$Bins_{i,q-1,t}$	**-0.01**	**-0.06**	**-0.02**
p-value	0.01	0.00	0.03
N	34	13	21
CD	-3.32	-3.02	
Av. Corr.	-0.02	-0.02	
BIC	5.84	5.87	
R-square	0.81	0.87	0.78
Hom.	4.69	0.96	1.78
Group 1:	14 22 29 37 38 41 42 56 70 73 84 93 96		
Group 2:	4 15 16 20 23 24 26 32 33 35 39 47 52 54 80 85 89 90 94 95 98		

Estimation results for the pooled FE models in Equation (8) (first column) and the grouped FE model in Equation (9) (second to fourth columns). N represents the number of forecasters, CD the cross-sectional dependence test of Pesaran (2004), *Av. Corr.* represents the average cross-sectional correlation of the model residuals, BIC is the modified BIC criterion proposed by Lin and Ng (2012), $R-square$ is the goodness-of-fit statistic, and *Hom.* represents the $\tilde{\Delta}$ dispersion test for parameter homogeneity proposed by Pesaran and Yamagata (2008). The CD and *Hom.* tests are standard normal distributed. The bottom of the Table provides the group membership based on the identifier provided in the Survey.