Are Professional Forecasters Bayesians?

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

We evaluate whether expectations of professional forecasters are consistent with the properties of Bayesian learning that claims that the expected uncertainty should decline with the forecast horizon and the weight used in updating the prior expectation should be bounded between zero and one. We construct a measure of uncertainty from the density forecasts of the Survey of Professional Forecasters and use it to measure the prior weight as the ratio of the precision in two consecutive quarters. Empirically, we find that these predictions are often violated, in particular when forecasters face a data release which is extreme relative to their prior density forecasts, and when large negative revisions occur, in particular for GDP. In addition, we find that the precision of the prior distribution is positively related to the prior weight for some forecasters, which suggests that they under-predict uncertainty in the first quarter of the year, and they later revise it upward in the second or third quarter. We also find that large surprises are related to lower absolute forecast error for approximately a quarter of the forecasters in our sample.

JEL Classification: E17, E37

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

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there is co-existence of agents that incorporate the most recent macroeconomic information while others persist using outdated forecasts.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). Coibion and Gorodnichenko (2010) and Andrade and Le Bihan (2013) provide empirical evidence on 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 about the way in which agents form expectations and their heterogeneity based on individual density forecasts rather than point forecasts. The source of density forecasts is the Survey of Professional Forecasters (SPF) that collects 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 SPF requires forecasters to provide density forecasts from the first to the forth quarter of a year with the objective to predict the outcome for that year. The target is thus fixed and it is possible to observe how individuals change their density forecasts in response to the incoming macroeconomic news that is released in a certain quarter. We conduct the analysis assuming that agents use a simple model of Bayesian learning to update their prior density forecasts. In particular, the hypothesis of normality of the prior and signal distributions produces a normal posterior distribution with mean equal to 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. The model also specifies the weight used to combine the prior mean and the signal which is obtained as the ratio of the prior and posterior precisions. Hence, the Bayesian learning model is capable to generate dispersion in beliefs through differences in prior expectations, and also as a result of heterogeneity in the weights used to update the prior. In particular, the latter channel occurs when forecasters use different models to interpret the incoming news, which leads to heterogeneous beliefs about its relevance to forecast the outcome. The objective is thus to investigate the contribution of differential interpretation to explain the evidence of heterogeneous beliefs without relying on restrictive identification assumptions.

Our empirical strategy consists of using the SPF density forecasts to construct an observable measure of the prior weight for each forecaster at each point in time. This is accomplished by extracting the variance of the densities which we then use to calculate the weight as the ratio of the prior and posterior precisions. The measured uncertainty and the weights can then be used to evaluate if the forecasts produced by professional forecasters are consistent with Bayesian learning and to what extent they are different across forecasters. We first evaluate empirically two testable implications of the learning model for the precision and weight measures: 1) the precision of the posterior distribution should not be smaller than the prior precision, and 2) the prior weight should be less or equal to 1 (and, of course, larger than 0). The empirical evidence shows frequent violations of these predictions. Forecasters often provide density forecasts for output growth and inflation that are more uncertain with regards to the forecasts that they produced in the previous quarter (for the same target). This is inconsistent with Bayesian learning since the additional macroeconomic data released in the current quarter should reduce their expected uncertainty if they consider the data informative or at least not increase when they regard the signal as uninformative. In those circumstances in which the posterior precision is smaller than the prior precision we find that the weight is larger than 1, which contradicts the second prediction of Bayesian learning. In addition, we find that the non-Bayesian behavior is common among most professional forecasters in our sample, although the frequency and magnitude of the deviations might be different. Furthermore, a fact that emerges from the analysis is that forecasters have disperse beliefs about the uncertainty of output growth and inflation which persist even in the last quarter of the year. We then investigate if the violations from Bayesian updating are purely random events or whether we can find a structural explanation for their occurrence. We construct a measure of individual surprise which captures the unexpected part of the GDP or GDP deflator latest release in the current quarter relative to the forecaster's prior distribution. We find that models for the prior weight that include the surprise variable are favored by most forecasters instead of a simple model with only quarterly dummy variables. The estimation results indicate that large surprises seem to be associated with higher weights and higher probability of non-Bayesian behavior, as argued in the theoretical models of Epstein (2006) and Ortoleva (2012). We also find that revisions of earlier released GDP and GDP deflator data lead some forecasters to increase the prior weight, although other forecasters interpret the revisions by assigning more weight to the new data rather than the prior expectation. In addition, another fact that emerges from the analysis is that for some forecasters the weight depends positively on the magnitude of the prior precision which suggests that they systematically under-estimate uncertainty in early quarters, and increase it later in the year. We also investigate if the surprise variable plays a role in determining the accuracy of the mean forecasts and find that low prior precision and large negative surprises lead to lower forecast errors for some forecasters.

This paper is organized as follows. In Section (2) we introduce the Bayesian Learning Model (BLM) and discuss its implications for the updating behavior of professional forecasters. In Section (3) we discuss the density forecasts provided by the Survey of Professional Forecasters (SPF) and in Section (4) we conduct an exploratory analysis of the empirical support for the BLM predictions followed by a regression analysis to understand the determinants of the observed non-Bayesian behavior of forecasters. Section (5) provides an empirical analysis of the relationship between forecast error and measures of surprises and, finally, Section (6) draws the conclusions of the paper.

2 Bayesian learning model

We discuss the Bayesian Learning Model (BLM) in the context of the timing of the information arrival and expectation formation of the SPF. 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 $q = \{1, 2, 3, 4\}$ indicates the quarter and $Y_{q,t-1}$ denotes the level of real GDP or GDP deflator in quarter q of year $t-1$. After observing the first release for Q4 of the previous year, the forecaster is asked to provide an expectation about the (year-on-year) growth rate of the variable in year t which we denote by y_t and is given by $y_t = (\bar{Y}_t - \bar{Y}_{t-1})/\bar{Y}_{t-1}$ and refer to the forecast of the probability density function of forecaster i by $f_1^i(y_t)$. Notice that as forecasters form an expectation of y_t in the first quarter they only observed the realization of the variable for the previous quarter due to the publication lag. Similarly, in the following quarters the forecaster observes the first data release for the previous quarter of that year, $Y_{q-1,t}$, and provides a density prediction, denoted by $f_q^i(y_t)$. In Q4 the forecaster has observed the data releases for quarter 1 through 3 and the only uncertainty remaining concerns the realization of the current quarter. The structure of the Survey is such that the forecast target y_t remains fixed throughout the year which allows to investigate how agents interpret the arrival of news when revising their expectations for the same target.

We assume that the density forecast of y_t in quarter $q-1$ (for $q = 2, 3$ and 4) of year t by individual i is normally distributed with mean $\mu^i_{q-1,t}$ and precision (inverse of the variance) equal to $\psi^i_{q-1,t}$. We interpret the density forecast in quarter $q-1$ as the agent prior expectation about the growth rate in the current year. In the following quarter, the forecaster observes the release of $Y_{q-1,t}$ and several other macroeconomic variables which can be interpreted as a generic signal $L_{q,t}$ about y_t defined as $L_{q,t} = y_t + \epsilon_{q,t}$ where $\epsilon_{q,t}$ represents an interpretation error of forecaster which we assume is distributed normally with mean $\nu^i_{q,t}$ and precision $\phi^i_{q,t}$. After observing 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

$$\mu^i_{q,t} = \rho^i_{q,t}\mu^i_{q-1,t} + (1-\rho^i_{q,t})(L_{q,t} - \nu^i_q) \tag{1}$$

$$\psi^i_{q,t} = \phi^i_{q,t} + \psi^i_{q-1,t} \tag{2}$$

where $\rho^i_{q,t} = \psi^i_{q-1,t}/(\phi^i_{q,t}+\psi^i_{q-1,t})$ 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 in Equation (2). The posterior mean in quarter q is thus given by a weighted average of the prior mean and the recently 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 assign the precision $\phi^i_{q-1,t}$ a value equal to zero so that $\rho^i_{q,t}$ will take a value of 1 and the posterior is thus anchored to the prior mean expectation. On the other hand, if the forecaster believes that the signal is very informative about y_t he/she will expect the signal to have high precision and thus $\rho^i_{q,t}$ will be close to 0 and a large fraction of the new information is incorporated in the posterior.

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 prior weight coefficients $\rho^i_{q,t}$ based on point predictions and assumes 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 the point forecasts to test their hypothesis about the learning model. Instead, in this paper we 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 moves from the first to the forth quarter, the individual precision of forecaster i in quarter q is given by

$$\psi^i_{q,t} = \psi^i_{1,t} + \sum_{j=2}^{q} \phi^i_{j,t} \tag{3}$$

for $q = \{2, 3, 4\}$. This Equation shows that the posterior precision of agent i is the sum of the prior precision in the first quarter and the cumulative precision of the signals. The BLM thus predicts that the posterior precision $\psi^i_{q,t}$ should not decrease as the target date gets closer since the ϕ^i_q are non-negative. This is a sensible prediction since uncertainty should reduce as we approach the target date and the forecaster should provide more precise density forecasts of the growth rate of the variable in the current year. Hence, for each forecaster it should hold that $\psi^i_{q,t} \geq \psi^i_{q-1,t}$ with equality holding only when the forecaster considers the latest signal totally uninformative. A related prediction that emerges from the BLM model concerns the prior weights $\rho^i_{q,t}$. Since the weight is the ratio of the prior and the posterior precision, the restriction in Equation (3) implies that

$$\rho^i_{q,t} = \frac{\psi^i_{q-1,t}}{\psi^i_{q,t}} \leq 1 \tag{4}$$

The weights $\rho^i_{q,t}$ are 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. Of course the weights are also restricted to be non-negative since they represent ratios of precisions (or variances).

The discussion above 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, and also by the fact that the prior weight used in revising their expectations should not be larger than 1. In other words, new information is likely to shift the center of the density forecasts but its dispersion should decrease as we approach the target date because forecasters should expect less uncertainty about the future outlook for output and inflation. The density forecasts provided by the SPF represent a unique dataset to test these hypotheses since they allow us to track the time evolution of the mean and variance of their forecasts. This approach is likely to provide more accurate insights on the expectation formation process and the hypothesis of differential interpretation relative to alternative approaches that treat the weight as one of several latent variables (Lahiri and Sheng, 2008, Patton and Timmermann, 2010 and Manzan, 2011).

3 Data

Forecasters contributing to the SPF provide their expectations about the probability that the year-over-year growth rate of real GDP and GDP deflator falls in ranges specified by the Survey that can be interpreted as a histogram forecast. The typical approach is to fit a parametric distribution to the histogram forecasts in order to derive the moments of interest. The two distributions that have been used in the literature are the normal distribution (see Giordani and Söderlind, 2003) and the beta distribution (see Engelberg *et al.*, 2009). In the first part of this Section we discuss the issues related to fitting these distributions to the individual histograms and then use the fitted distributions to analyze the empirical validity of the predictions of the BLM discussed above. Our results indicate that there are significant departures from "Bayesianity" which we then try to explain empirically.

3.1 Density forecasts in the SPF

The American Statistical Association (ASA) and the National Bureau of Economic Research (NBER) started to collect expectations from professional forecasters about macroeconomic variables in 1968 and since 1992 the Survey is administered by the Federal Reserve Bank of Philadelphia. SPF collects expectations of forecasters employed in the private sector and research organizations about several macroeconomic variables with the forecast horizon ranging from the previous quarter (formed after the first release is available), the current quarter (which is released next quarter), and up to 4 quarters ahead. In addition, the SPF asks participants to provide their expectation about the probability that average year-over-year growth rates of current and next year GDP and GDP deflator fall in specific value ranges. This implies that for the current year there are four density forecasts available and eight for the two year forecasts. Croushore (1993) provides a detailed description and discussion of the SPF. More recently, the Bank of England has started a Survey of External Forecasters with the aim to collect both point and density forecasts (see Boero *et al.*, 2008 and 2012, for a detailed analysis).

The SPF density forecasts are special among surveys because they provide a measure of the mean/median outcome expected by forecasters and, quite uniquely, an individual measure of the expected uncertainty about the future growth in output and prices. 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 given the common practice of considering them equivalent although conceptually different. They measure uncertainty by the standard deviation of the aggregate density forecast obtained by averaging the individual forecasts. They found that disagreement and uncertainty have positive and high correlation which provides support to the use of measures of forecast disagreement as a 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. The SPF is an invaluable tool to investigate this issue because it provides an observable measure of uncertainty at the individual and aggregate level. 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 (2013) 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 (3) 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, and Clements, 2010) 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.

3.2 From histograms to continuous distributions

The SPF provides participants with a set of intervals and asks them to assign probabilities to the event that the growth rate for the current year, y_t, will fall in each of these intervals. Denote by $p^i_{j,q,t}$ the probability forecast in quarter q of year t that individual i assigns to interval j

(for $j = 1, \cdots, J$) that has left and right bounds given by x_{j-1} and x_j, respectively. Fitting a parametric density to the histogram requires minimizing the following quantity:

$$\min_{\theta^i_{q,t}} \sum_{j=1}^{J} \left[\left(\sum_{k=1}^{j} p^i_{k,q,t} \right) - F(\bar{x}_j, \theta^i_{q,t}) \right]^2$$

where $\bar{x}_j$ represents the mid-point of the interval and $\sum_{k=1}^{j} p^i_{k,q,t}$ is the cumulative probability provided by the forecaster in the SPF up to interval j and $F(\cdot, \theta^i_{q,t})$ represents the CDF of the parametric distribution which depends on a vector of parameters $\theta^i_{q,t}$. Giordani and Söderlind (2003) assume that $F(\cdot)$ is the normal distribution which is convenient due to its simplicity and ease of interpretability. However, Engelberg *et al.* (2009) argue that the normal distribution might not be an appropriate assumption for the SPF density forecasts because it requires symmetry of the distribution which is at odds with the empirical evidence that in many instances forecasters 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.

In this paper we approximate the histograms using the normal distribution. One reason for this choice is to have a consistent distributional assumption in the theoretical model and in the empirical analysis. Although it is certainly true that occasionally forecasters provide asymmetric histograms, it does not seem a widespread phenomena across forecasters and across time. Overall, the assumption of normality of the SPF seems an appropriate distribution to use, although in some cases it might be at odds with the evidence. Empirically, we find that the correlation between the fitted mean of the normal and beta distribution is 0.9998, and the correlation between the standard deviations is 0.9934 with the average difference between the implied means of -0.0034 and for the standard deviations the difference is 0.068. For the purpose of this paper we feel that the consistency between the theoretical and empirical models is an important consideration and thus adopt the assumption of normality, although it might come at the cost of the misspecification of the density predictions of some forecasters in some quarters.

4 Application

Every quarter we fit a normal distribution to the histogram forecast of each individual as described above, and obtain parameter estimates $\theta^i_{q,t} = [\mu^i_{q,t}, 1/\psi^i_{q,t}]$ which represent the mean and precision of the density forecast for forecaster i in quarter q of year t. In the empirical analysis we consider the current year forecasts, and include in our sample only those forecasters that provide a minimum of 30 two-quarter consecutive predictions over the period considered. In addition, although the SPF density forecasts are available since 1968 there are several changes of definition of the output measures from GNP to GDP, and from nominal to real, which makes the analysis over time challenging. We thus start the analysis in 1982 for both real GDP and GDP deflator, and ending in Q4 of 2012. This gives us a sample of 23 forecasters for PGDP and 24 for GDP over 124 quarters.

The first prediction that we want to test is whether the precision of the density forecasts is increasing as the target date approaches. Figure (1) shows the (time) average precision of each forecaster for density forecasts provided in quarter 1 through 4. Overall, the precision increases toward the end of the year and there is significant dispersion of the precision measure among forecasters. For example, in the first quarter some forecasters expect a precision of 1 while others expect it as large as 8. The dispersion of precision seems to increase in later quarters since some forecasters increase the precision of their posterior more rapidly than others. However, a closer look at the graph shows that some of these lines do not monotonically increase over time, which is inconsistent with the prediction of the BLM model that the precision of the density forecasts should be non-decreasing. Another way to evaluate the validity of this prediction is to plot the (time) average prior weight $\rho^i_{q,t}$ given in Equation (5). The prior weight is defined as the ratio of the prior and posterior precisions, and should thus be bounded between 0 and 1. Figure (2) shows the average weight from the second to the forth quarter with each line representing a forecaster. It is clear that there are several forecasters that have weights larger than 1, in particular in the second and third quarters and, in some cases, also in the forth quarter. The finding that some forecasters have (average) prior weights above 1 in some quarters indicates that the precision of their posterior density forecasts decline relative to the prior precision. The inconsistency with the BLM model arises because the newly released signal should convey information, and thus increase precision relative to the previous quarter or, if the signal is expected to be uninformative, the prior and posterior precision should be equal, and the weight should be one. However, the evidence seems to suggest that some forecasters form density predictions that underestimate future uncertainty, and in later quarters they revise their distribution forecast by increasing its dispersion. This inconsistency seems more apparent when plotting the $\bar{\rho}^i_q$ in Figure (2) rather than the average $\bar{\psi}^i_q$ in Figure (1). This is due to the fact that averaging the precision forecast misses the dynamic aspect of taking the ratio of the prior and posterior precisions, i.e., $\sum_{t=1}^{T} \psi^i_{q-1,t}/\psi^i_{q,t}$.

In order to analyze in more detail this finding, we divide the forecasters in two groups: a group which we call Bayesians Forecasters (BF) is composed of those forecasters with $\bar{\rho}^i_q$ smaller than 1 at all quarters, and the remaining forecasters are pooled in a group that we denote as Non-Bayesians Forecasters (NBF), since they violate the prediction of the BLM in at least one quarter. In practice, we use a threshold of 1.05 rather than 1 to avoid the fact that small errors in fitting the parametric distributions might lead to spuriously classify a forecaster in one group rather than the other.

Figure (3) shows the average prior weight $\bar{\rho}^i_q$ by quarters, and separately for BF and NBF. When considering PGDP there are 7 forecasters out of 23 that provide forecasts consistent with BLM, and 16 that are not consistent. For GDP we have 24 forecasters who are equally divided between BF and NBF. The two left plots of Figure (3) show that forecasters that are classified as Bayesians have weights close to 1, and declining very slowly in the second and third quarter, which indicates that (on average) they put a small weight on the signal contained in the new data release while they prefer to hold on to their prior expectation of uncertainty. However, in the forth quarter they seem to increase their posterior precision, and assign a lower weight to the prior, which oscillates between 0.5 and 0.75 across forecasters. The average prior weights for the NBF are shown on the two right plots and indicate different patterns across quarters. Some forecasters seem to overestimate the precision in the first quarter which they later reverse by revising the precision downward. This pattern of weights larger than 1 in the second quarter and smaller than 1 in the last two quarters of the year seems to be more common for density forecasts of real GDP rather than PGDP since it characterizes the behavior of 5 out of 16 NBF. Another type of revision pattern that emerges from the plot for PGDP consists of 6 forecasters with weight smaller than 1 in the second and forth quarter, but larger than 1 in quarter 3. This pattern is the result of a decrease in expected uncertainty in quarter 2 relative to quarter 1 followed by an increase in the following quarter, and again a decrease in the last quarter of the year. One possible rationale for this behavior is that forecasters consider the third quarter as the crucial time to evaluate the available information and their density forecasts, which might then trigger the forecaster to reconsider their expectation of uncertainty.

So far we have discussed the characteristics of the individual weights which have been obtained by averaging over time the ratio of prior to posterior precisions. However, the prediction of the BLM should actually hold at every point in time rather than on average, and the question is then to evaluate if forecasters incorporate this restriction anytime they produce a density forecast. To investigate this issue we calculate for each forecaster i the fraction of $\rho^i_{q,t}$ that are larger than 1.05 from the second to the forth quarter and the results are shown in Figure (4) for BF (left plots) and NBF (right). The fact that BF on average satisfies the prediction of the BLM model hides the fact that they occasionally violate it, in some cases as often as in 25% of the forecasts submitted, in particular in the second and third quarter. However, in the last quarter the fraction of inconsistent revisions tends to be smaller for most of the Bayesian forecasters. On the other hand, it is not surprising to find that for some NBF the fraction of inconsistent revisions exceeds 50%. Also for this group we find that the majority of deviations happens in the second and third quarter and we can identify some of the same patterns for the average weights that were discussed above. Another relevant issue relates to the magnitude of the deviations of the weights from the theoretical value of 1 when the threshold is exceeded. We answer this question in Figure (5), and the results clearly show that deviations for BF are relatively smaller compared to NBF for which, in particular in quarter 2 and 3, the prior precision is as much as twice or three times the posterior precision.

4.1 An empirical model of the prior weight $\rho^i_{q,t}$

The main finding of the previous Section is that a relevant fraction of forecasters in the SPF deviates from proper Bayesian updating by increasing the variance (i.e., lowering precision) when revising their density forecasts. As discussed above, this behavior is inconsistent with the prediction of the BLM, since forecasters should expect a decline in uncertainty as the target date approaches. Different explanations could be argued for this behavior. One is that forecasters lack incentives to provide accurate density forecasts because these forecasts are not subject to the careful scrutiny of clients and media who are mostly interested in point forecasts. They might thus be very attentive in crafting a prediction for the mean of the density, but guesstimate the magnitude of its precision. Stark (2013) conducted a Survey of SPF participants and found that only 8 forecasters use the density forecasts in their analysis while 17 forecasters produce the forecasts only for the SPF. In addition, 11 participants declare to use the results of the survey's density forecasts in their work as opposed to 15 who do not. This suggests that the findings in the previous Section of weights larger than 1 might be the outcome of inattention when it comes to predict the dispersion of the forecast distribution. A more structural explanation for non-Bayesian behavior is provided by Epstein (2006) and Epstein *et al.* (2008, 2010). They argue that non-Bayesian updating might arise in response to a signal that is interpreted as positive or negative by forecasters, and leads them to produce a posterior forecast which is inconsistent with the 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. In the context of the SPF, it could be the case that in the second quarter the forecaster realizes that the signal is at odds with the first quarter prior density, and thus produces a new prior which is then updated to incorporate the newly released information. Since we do not observe the revised prior belief of forecasters, we might thus conclude that the posterior is inconsistent with the prior, although it might be the outcome of a retroactive change in prior beliefs of the forecaster. Another explanation that has been offered in the literature suggests that forecasters might re-assess their prior forecasts in response to signals that have low probability to occur based on their prior forecasts (Ortoleva, 2012). In this case, a forecaster might interpret the fact that the realization of the signal (e.g., the quarterly release of the variable) is not very likely to happen based on her prior density as an indication of misspecification of the forecasting model which leads to re-evaluate her beliefs. These theoretical models thus assume that the signal triggers a re-assessment of the forecaster's beliefs and leads to posterior densities that are inconsistent, in a Bayesian sense, with their reported priors in the previous quarter. In practice, the information available in the SPF might not be sufficient to be able to distinguish between the two explanations. In the empirical analysis described below we pursue the more modest goal of investigating the variables that explain the dynamics across quarters and forecasters of the prior weight, and the role played by a measure of news or surprise in explaining the non-Bayesian behavior.

4.2 Empirical specification

We assume the following model for the prior weight of forecaster i in quarter q of year t, denoted by $\rho^i_{q,t}$:

$$\rho^i_{q,t} = \beta_{i,2}Q2_{q,t} + \beta_{i,3}Q3_{q,t} + \beta_{i,4}Q4_{q,t} + X^{'}_{i,q,t}\gamma_i + \epsilon^i_{q,t} \tag{5}$$

where the βs and γ are forecaster-specific parameters to be estimated, and the variables $Q2_{q,t}$, $Q3_{q,t}$, and $Q4_{q,t}$ represent quarterly dummy variables that take value 1 if quarter q is the second, third, or forth, respectively, and are equal to zero otherwise. The vector $X^i_{q,t}$ includes forecaster-specific variables that we use to explain the variation over time of the prior weight. In particular, we include the previous value of the weight (in quarter 3 and 4) to evaluate if there is persistence or reversal in the dynamics of the weight and whether BF and NBF behave differently. Another variable that we include is the level of the precision in the previous quarter, $\psi^i_{q-1,t}$. The reason for including this variables is to assess if forecasters with high precision are more likely to reduce or increase their prior weight in the following period than forecasters that expected lower precision. In addition, we create a surprise variable that aims at measuring the unexpected component of the latest data release which might be responsible for triggering a (non-) Bayesian update of the prior weight as argued by Epstein (2006) and Ortoleva (2012) among others. We discuss the construction of this variable below.

We define the surprise for forecaster i in quarter q of year t, denoted by $S^i_{q,t}$, as follows:

$$S^i_{q,t} = \sqrt{\psi^i_{q-1}}\left(\tilde{\mu}^i_{q-1,t} - \mu^i_{q-1,t}\right) \tag{6}$$

where $\mu^i_{q-1,t}$ and $\psi^i_{q-1,t}$ are the mean and precision of the density forecasts of individual i produced in the previous quarter and $\tilde{\mu}^i_{q-1,t}$ (for $q \geq 2$) represents the expected value of the growth rate of PGDP or GDP obtained by replacing the nowcast for quarter $q-1$ (i.e., the point forecast $E^i_{q-1,t}(Y_{q-1,t})$) with the first release available in quarter q (i.e, $Y_{q-1,t}$). More specifically, we construct $\tilde{\mu}^i_{q-1,t}$ as follows:

$$\tilde{\mu}^i_{q-1,t} = \frac{1}{4}\left(\sum_{i=1}^{q-1} Y_{i,t} + \sum_{j=q}^{4} E^i_{q-1}(Y_{j,t})\right)/\bar{Y}_{t-1} \tag{7}$$

where the denominator $\bar{Y}_{t-1}$ represents the average level of the variable in the previous year calculated using the data vintage available in quarter $q-1$. Instead, the numerator represents the average of the data releases available in the quarter $q-1$ vintage, the realization of the previous quarter (which is released in quarter q), and the point forecasts for the current and future quarters produced in quarter $q-1$. We use the point forecasts because they are formulated in the level of the variable, and they allow to calculate the year-over-year growth rate in a given quarter. In this way we transform the recent data release from the level of the variable to a growth rate which can be related to the prior density forecasts to define a surprise measure. We define the surprise $S^i_{q,t}$ as the difference between the updated mean from the prior mean standardized by the predicted uncertainty of forecaster i in the previous quarter. Hence, the measure captures the strength of the news in the most recent data release relative to the prior view of the forecaster.

In addition, we consider the possibility that the prior weight $\rho_{i,q,t}$ might be non-linearly related to the surprise variable, such as the case that forecasters react differently to a positive or negative surprise. We thus create the variables $S^+_{i,q,t}$ and $S^-_{i,q,t}$ which are equal to the surprise when the surprise is positive and negative, respectively, and zero otherwise. Another argument put forward earlier is that forecasters might change the prior weight in response to unlikely surprises. We thus define as unlikely the bottom and top 5% of each forecaster surprise distribution and define the dummy variables $I^{highS}_{i,q,t}$ and $I^{lowS}_{i,q,t}$ that take value one if $S^i_{q,t}$ is smaller (larger) than the forecaster's 0.05 (0.95) quantile, and are equal to 0 otherwise.

In Figure (6) we show a histogram of the surprise for PGDP and GDP obtained by pooling the surprises over time and across forecasters. The distribution appears to be very peaked at the center (close to zero) and asymmetric due to the long left tail for PGDP and right tail of GDP. The asymmetry in the tails suggests that, when large surprises occur, forecasters are more likely to overestimate the inflation rate while for output they seem to underpredict its value. We also calculate the probability to observe a growth rate equal or smaller than $\tilde{\mu}^i_{q-1,t}$ based on the density forecast produced by each individual in quarter $q-1$. Given our assumption of normality the probability is calculated as $\Phi^i_{q-1,t}(\tilde{\mu}^i_{q-1,t})$, with $\Phi^i_{q-1,t}(\cdot)$ denoting the prior normal CDF with mean $\mu^i_{q-1,t}$ and variance $1/\psi^i_{q-1,t}$. In Figure (7) we find that for both PGDP and GDP the probabilities are concentrated in the interval 0.4 to 0.6, which suggests that most often the forecasters are not surprised by the recent announcement of the data. Furthermore, the peaks for small p-values in the PGDP histogram and for large values for GDP show, similarly to Figure (6), that the news content of data releases are occasionally considered extremely unlikely by forecasters based on their forecasts of the previous period.

Another source of unexpected news for forecasters might come from the announcement of revisions of data released in previous quarters. Statistical agencies provide every quarter the first release for the previous quarter (i.e., $Y_{q,t}$) and also revisions of earlier released data for the current and previous years. Notice that in Equation (6) we defined the surprise based on the vintage of data available in quarter $q-1$ (except for the first release of $q-1$). However, forecasters also observe the new vintage of data and these data revisions might be substantial, and may induce forecasters to change the weight they assigned to their prior and to the signal. We define the revision in quarter q of year t for forecaster i as follows:

$$R_{i,q,t} = \hat{\mu}^i_{q-1,t} - \tilde{\mu}^i_{q-1,t} \tag{8}$$

where $\hat{\mu}^i_{q-1,t}$ is constructed in the same way as $\tilde{E}^i_{q-1}(y_t)$ in Equation (7) using the latest available vintage released in quarter q instead of the previous quarter vintage. $R_{i,q,t}$ can thus be interpreted as the surprise content of the data release which is due to revision of earlier data. For example, in the last quarter of the year the revision $R_{i,4,t}$ might be different from zero if there were revisions to the level of the variable for the first and second quarter of the current year, or revisions about the previous year which affect the denominator of the growth rate. In the regression analysis for the prior weight $\rho^i_{q,t}$ we decided to include the revision as a dummy variable that equals 1 in case of a large positive or negative revision. We define as $I^{highR}_{i,q,t}$ the dummy variables that take value 1 if the revision is larger than the 0.95 quantile of forecaster's i revision distribution and by $I^{lowR}_{i,q,t}$ if the revision is smaller than the 0.05 quantile. In Figure (8) we show the histogram of the revisions for PGDP and GDP pooled across forecasters and time. The revisions for PGDP appear to be smaller in magnitude and more concentrated relative to those for GDP, which can be as large as 1%, and in some cases even 2%. This is expected since revisions to GDP data releases are typically more biased and volatile relative to the PGDP ones (Aruoba, 2008).

4.3 Estimation results

We consider several specifications of the model for the prior weight in Equation (5) which differ for the set of predictors that are included in the vector $X^i_{q,t}$. We estimate the model individually for each forecaster by OLS although it is possible that the errors are correlated across forecasters due to unobserved common shocks which causes OLS to be inconsistent. Hence, after estimating the models by OLS we test for the null hypothesis of no cross-sectional dependence and, in case of a rejection, we re-estimate the model using the Common Correlated Estimator (CCE) proposed by Pesaran (2006), which provides consistent estimates even in the presence of unknown common shocks. We 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

$$CD = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{i,j}} \hat{\xi}_{i,j} \right) \tag{9}$$

where $\xi_{i,j}$ represents the correlation coefficient of the errors in Equation (5) 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 that follows, we found no evidence of dependence in the forecaster errors for PGDP and thus estimate the model by OLS. Instead, the residuals for the GDP regression show significant evidence of correlation across forecasters, and we thus report the CCE estimates.

Table (1) and (2) provide the estimation results for Equation (5) with Model A indicating the model that includes only the quarterly dummy variables and Model B to G including additional variables, such as the lagged value of the weight, the prior precision, the surprise variable in different functional forms, and the revision dummy variables. For each variable we provide the Mean Group (MG) estimator along with the number of forecasters for which the coefficient is significant at 10% level and report the results for all forecasters and also separately for BF and NBF (in this case the MG is calculated on the subset of forecasters). At the bottom of the Tables we provide summary statistics such as the average cross-correlation coefficient of the residuals and the CD test statistic discussed above (which is standard normally distributed), in addition to the mean/min/max of the R^2 of the individual regressions and the mean R^2 for the BF and NBF groups. Finally, we report the number of forecasters for which each model is selected based on the AIC and BIC selection criteria.

The results for the GDP deflator in Table (1) show that the lagged prior weight is a significant variable for several forecasters and the MG estimate is in all cases negative and of similar magnitude for both BF and NBF. This suggests that the prior weight for some forecasters alternates between high and low values as time progresses. The prior precision seems an important variable for NBF (between 3 and 6 forecasters out of 15) and for some BF (1-3 forecasters out of 8). The positive sign of the coefficient indicates that those forecasters who have high prior precision (low variance) in a certain quarter are likely to assign a high weight (possibly larger than 1) to the prior when updating their forecast. This is consistent with our earlier evidence that some individuals produce narrow density forecasts in the early quarters of the year which they later revise by increasing the variance (reducing the precision) of the forecast. This behavior leads to weights that are larger than one since they are obtained by taking the ratio of the prior precision to the posterior one. Furthermore, the estimation results for Model D show that the surprise variable $S_{i,q,t}$ is significant for 8 forecasters, 6 of which are NBF. Interestingly, the negative sign of the MG estimator for NBF suggests that these forecasters seem to have high weights when surprises are negative and low weights when surprises are positive. A one standard deviation change in the surprise variable leads to a change of the weight by -0.17. This evidence suggests that these forecasters weight more the real-time information when they experience a positive surprise, while maintaining the posterior close to the prior mean when affected by a negative surprise. To better assess the dependence of the prior weight on the surprise and their possible asymmetric relationship, in Model E we consider separately positive and negative surprises. For this model we find that NBF forecasters seem to have higher weights as the surprise becomes larger on either sides (negative coefficient for $S^{-}_{i,q,t}$ and positive for $S^{+}_{i,q,t}$). This is also consistent with the findings for Model F and G in which we include the surprise and two dummy variables that take value 1 when the surprise is large either on the left or right side of the surprise distribution. The results show that the coefficients of these dummy variables are significant for one third of the NBF and both have positive signs, although the coefficient for large negative surprises is much larger relative to the one for large positive surprises. Finally, in Model G we also include dummy variables for large positive and negative revisions and find that forecasters do not seem to adjust the weight in response to this type of surprises. In terms of the best performing model across forecasters, the AIC and BIC criteria provide different answers, which is expected since they use different penalizations for the complexity of the model. AIC indicates that models that include the surprise are preferable, and it selects models D and F for about half of the forecasters. On the other hand, BIC imposes the harshest penalization, and it is thus not surprising to find that it favors smaller models such as Model A (which includes only the quarter dummy variables) for 8 forecasters and models D and F (4 forecasters). The fact that no model seems to provide the best fit for a large fraction of forecasters is probably related to the intrinsic heterogeneity in the way forecasters update their density forecasts in response to new information which has been documented in the previous Section. However, the different selection criteria seem to agree overall that the surprise variable provides a better characterization of the prior weights.

Table (2) shows the regression results for the prior weight obtained from the GDP density forecasts. Most of the findings discussed for PGDP are confirmed by the analysis of GDP in terms of magnitude and sign of the coefficient estimates, and in terms of the significance across forecasters.

However, we find that in Models F and G forecasters seem to have a different reaction to large surprises with some BF more likely to increase their weight on the prior and NBF doing the opposite, in particular for large negative surprises. Another difference is represented by the role of the revision dummy variable. The results for GDP indicate that 4 forecasters (2 BF and 2 NBF) respond to large negative revisions and only two forecasters react to positive ones. In addition, we find that the MG estimator for large negative revisions is negative for BF and positive for NBF which suggests that these surprises lead BF to give more weight to the real time information while NBF do the opposite. In terms of model selection, the AIC and BIC criteria indicate that models D to G are selected by the vast majority of forecasters, which points to the relevance of the surprise variable (in the different functional forms considered) in explaining the learning and updating process of professional forecasters.

4.4 Binary model

In the previous Section we consider an empirical model for the prior weight $\rho_{i,q,t}$, and investigated the variables that could explain its variation over time and across individuals. However, the analysis focused on explaining the magnitude of the weight which offers a limited perspective on the factors behind the non-Bayesian behavior of forecasters. We thus perform an analysis similar to the previous Section, but considering as dependent variable the binary event that takes value 1 if $\rho^i_{q,t}$ is larger than 1 and 0 otherwise. We use a probit model and consider the same variables and specifications that we adopted in the previous Section. The model is estimated individually for each forecaster by maximum likelihood and in Tables (3) and (4) we report the average estimates along with the number of forecasters for which the coefficient is significant. We also test for the presence of cross-sectional dependence using the CD test discussed above, which is appropriate also for residual measures from limited dependent variable models as argued by Hsiao *et al.* (2012). For each specification we also report the number of forecasters for whom the model is best according to AIC and BIC as well as a measure of fit given by the fraction of correct predictions.

The results for PGDP in Table (3) show that the probability of non-Bayesian behavior (i.e., $\rho_{i,q,t} > 1$) increases when forecasters experience a large positive surprise, both for BF and NBF. In other words, some forecasters increase the expected uncertainty relative to their belief in the previous quarter when the latest PGDP release is significantly higher than they expected. This is consistent with the earlier discussion, although we find that large negative news do not increase the probability that the weight is larger than one. All other variables have a negative impact on the prior weight, and the lagged prior weight and the surprise variable are significant for a large fraction of forecasters. On the other hand, the prior precision and the revision dummy variables are significant for few forecasters, which indicates that they might not be relevant contributors to the non-Bayesian behavior of forecasters. For all models the CD test statistic is calculated using the generalized residuals for the probit model, and it is not significant at conventional significance levels. The AIC and BIC indicate that for a large fraction of forecasters the simple model with just the quarter dummy variables outperforms all other specifications. For the remaining forecasters, the most often chosen specifications are G, D and E. These findings seem also to hold for GDP in Table (4) with the only difference that also the dummy for large negative revisions is significant for 4 forecasters (3 of which are NBF), and with a positive coefficient. This suggests that some forecasters might be surprised by large unexpected downward revisions to the earlier released data, and they might respond by increasing the variance (or decreasing the precision) of their density forecasts relative to their forecast of the previous quarter. Similarly to the linear regression case, we find that the CD test statistic for the GDP probit model rejects the null hypothesis of no cross-sectional dependence in the errors among forecasters. This suggests that the results provided in the Table for GDP should be interpreted cautiously, although the magnitude of the correlation, which ranges between 0.053 and 0.067, is quite modest to expect dramatic changes in these findings once dependence across forecasters is taken into account.

5 Are Bayesian forecasters more accurate?

In this Section we investigate if the non-Bayesian updating adopted by some forecasters has an effect on their forecast accuracy. In Figure (9) we show the box plot of the absolute forecast error in quarter 1 through 4, where we calculate the error as the difference between the realization of the growth rate in year t and the mean of the fitted normal density forecast. As expected, the median absolute error reduces later in the year, with only few differences between BF and NBF. For PGDP, the NBF seem to experience a reduction of error only in the third quarter, while for BF the reduction starts already in the second quarter. In the forth quarter, NBF have a median error which is slightly larger relative to BF, while for GDP we find the opposite result that the median absolute forecast error for NBF is smaller than for BF. For both variables, the results indicate that forecasters make errors as large as 0.5% and with the third quartile close to 0.4% in the forth quarter, which is quite large considering that in the following quarter they are able to calculate y_t. An alternative way to relate the realization y_t to the density forecast is to calculate the Probability Integral Transform (PIT) (see Diebold *et al.*, 1997) which is defined as $\Phi_{i,q,t}(y_t)$, with $\Phi_{i,q,t}(\cdot)$ denoting the normal CDF forecast of individual i in quarter q of year t. If the individual is using the correct model to produce forecasts, the PITs should be uniformly distributed in the unit interval, with deviations from uniformity indicating the misspecification of the density forecasts.

In Figure (10) we show the box plot of the PIT by quarter for PGDP and GDP, and separately for BF and NBF, along with horizontal lines at 0.25, 0.5, and 0.75 to facilitate the visual evaluation of the uniformity assumption. For GDP (bottom two graphs) we find that the median PIT is quite close to 0.5 although the interquartile range appears to be wider than expected in the first quarter and smaller in the forth quarter (in particular for NBF). This indicates that professional forecasters under-estimate the probability of tail events at the beginning of the year (in particular for large positive events), but overestimate the likelihood of the extremes at the shortest forecast horizon. This could be related to the earlier discussion that in the first quarter some forecasters produce density forecasts with a small variance (high precision) which they later revise upward (downward). Instead, for PGDP we find that the median PIT is typically below 0.5 and the interquartile range is shifted downward relative to the 0.25-0.75 interval, in particular from the second to the forth quarters. This suggests that, for most forecasters and quarters, the realization of the GDP deflator happens on the left tail of the density forecasts provided by the SPF participants. This conclusion relates also to the earlier discussion of the surprise histogram for PGDP that showed that many forecasters expect inflation to be higher than it turns out to be. These two facts together point to the fact that, despite the negative surprises about inflation, forecasters overweight their prior beliefs of higher inflation relative to the real-time information so that the realized inflation rate occurs often on the left tail of density forecasts.

To better understand the determinants of the forecast errors, we estimate a regression model using the same independent variables adopted in the previous Section to explain the prior weight. Notice that all these variables are available to the forecasters once the statistical agencies release the data for the previous quarter. On the other hand, the dependent variable represents the ex-post accuracy of the density forecasts and there is thus a clear causality between the RHS and the dependent variable. Among the regressors, we are particularly interested in evaluating whether the surprise variable discussed in the previous Section has any explanatory power for the forecast error, which would indicate that forecasters might have over or under-reacted to the informational content of the latest data release. The regression model is given by

$$|e_{i,q,t}| = \beta_{i,1}Q1_{q,t} + \beta_{i,2}Q2_{q,t} + \beta_{i,3}Q3_{q,t} + \beta_{i,4}Q4_{q,t} + X^{'}_{i,q,t}\gamma_i + \epsilon_{i,q,t} \tag{10}$$

where the error $e_{i,q,t}$ is defined as the difference between the realization y_t and the mean of the density forecast $E_{i,q}(y_t)$. In the vector $X_{i,q,t}$ we include the same variables that were considered in the previous Section, except that the lagged dependent variable is now represented by the absolute error of the previous quarter $|e_{i,q-1,t}|$. The application of the CD test to the OLS residuals indicates the presence of statistically significant cross-sectional dependence so we estimate the model using the CCE estimator of Pesaran (2006) for both variables and the results are reported in Table (5) for PGDP and Table (6) for GDP. The AIC and BIC criteria indicate that Models B, D to E are selected by a large fraction of forecasters. The difference in results between the two criteria is partly due to the higher penalization assumed in BIC relative to AIC, but also because the CCE estimator increases significantly the number of parameters by requiring to include the cross-sectional averages for every additional variable included in the model.

In terms of the relevant variables for PGDP, we find that the coefficient estimates of the lagged absolute error are positive and statistically significant at 10% for approximately half of the forecasters, and in particular for NBF, who seem to have higher error persistence in comparison to BF. The prior precision is significant for only a few forecasters, and the positive estimates indicate that forecasters with larger prior precision (smaller variance) are likely to be characterized by higher absolute forecast errors. We also find that the surprise variable is more significant when it is considered in a nonlinear form. Model E indicates that some forecasters have smaller forecast errors following large surprises (either positive or negative) than small surprises (relative to the median). This could be interpreted as evidence that these forecasters might be more attentive to the real-time information and to the formation of their density forecasts when faced with large unexpected news (with regards to their prior forecast) which thus leads to smaller absolute forecast errors. The results for Model F provide a similar interpretation since large negative surprises (5% lowest surprises of each forecaster) have a MG estimate which is negative and is thus associated with lower absolute errors. This variable is significant for 4 NBF. On the other hand, the large positive surprise variable is only significant for one forecaster and the MG estimate is very small, which suggests that this type of surprise might not be relevant to explain the performance of forecasters. Contrary to the case of PGDP, we find that large revisions, in particular negative ones, have an effect on the accuracy of density forecasts for GDP growth. In particular, we find that after a large negative revision to the earlier released data some BF experience an increase in absolute error, while some NBF a decline in the measure. Another difference between the results for GDP and PGDP is that the residuals are characterized by stronger correlation across forecasters. For Model A the correlation estimate is 0.486 and the CD test statistic is over 50. The application of the CCE estimator in Models B to H reduces significantly the cross-correlation to a range between 0.035 and 0.051 (in absolute value), and the CD statistic takes values between 2 and 3. Although the CD statistic still rejects the null of no cross-correlation in the residuals at 5%, the size of the correlation is not so large to expect a relevant change in the results. Overall, these results indicate that the more relevant predictor of future performance is the past performance of the forecaster, although measures of surprises and revisions are able to capture some of rthe variability over time of the absolute forecast error.

6 Conclusion

Are professional forecasters Bayesians? We find empirical evidence that they form expectations that occasionally are inconsistent with the predictions of a model of Bayesian learning. The first prediction is that forecasters should expect uncertainty to decline or stay constant nearing the target date while the second prediction is that the weight used to update the prior expectation should be smaller or equal to 1. Our empirical evidence of violations of these predictions is based on measures of individual uncertainty and prior weight that we construct from density forecasts, rather than relying on point forecasts as in the recent literature on testing models of expectation formation. The reasons for these violations is that forecasters sometimes report a density forecast with larger variance relative to the forecast they provided in the previous quarter. This fact goes against the principle that the availability of more information should not increase the expected uncertainty of the revised forecast. Furthermore, we also find that professional forecasters are significantly heterogeneous in the way they process information and respond to macroeconomic surprises.

We investigate empirically the possible determinants of this behavior and find that, for some forecasters, a measure of surprise or unexpected news contributes to explaining the time-variation of the prior weight, and the non-Bayesian behavior of professional forecasters. We find that forecasters are more likely to update their prior in a non-Bayesian manner when hit by a news release that is unlikely relative to their prior density forecasts. It seems also that large negative revisions to GDP prompt some forecasters to increase (relative to the prior) their expected uncertainty, while this is not the case for PGDP inflation. Another relevant variable to explain the variation in weights is the precision of the prior distribution, which is significantly and positively related to the prior weight for several forecasters. We also find that large surprises are related to lower absolute forecast error for approximately a quarter of the forecasters in our sample. This is probably due to the fact that a large unexpected data release motivates agents to re-evaluate their prior forecast, and revise it by increasing the expected variance. This might lead to more realistic forecasts and to higher ex-post accuracy.

References

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

Aruoba, S Borağan (2008). Data revisions are not well behaved. *Journal of Money, Credit and Banking*, **40,** 319–340.

Boero, Gianna, Smith, Jeremy and Wallis, Kenneth F (2008). Uncertainty and disagreement in economic prediction: The Bank of England Survey of External Forecasters. *Economic Journal*, **118,** 1107–1127.

Boero, Gianna, Smith, Jeremy and Wallis, Kenneth F (2012). The measurement and characteristics of Professional Forecasters' uncertainty. *working paper*.

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

Clements, Michael P (2013). Forecast uncertainty ex ante and ex post: US inflation and output growth. *Journal of Business & Economic Statistics*, forthcoming.

Coibion, Olivier and Gorodnichenko, Yuriy (2010). Information rigidity and the expectations formation process: A simple framework and new facts. *working paper*.

Croushore, Dean (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, **11/12,** 3–13.

Diebold, Francis X, Tay, Anthony S and Wallis, Kenneth F (1997). Evaluating density forecasts of inflation: the Survey of Professional Forecasters. Technical Report. National Bureau of Economic Research.

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**.

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

Hsiao, Cheng, Pesaran, M Hashem and Pick, Andreas (2012). Diagnostic tests of cross-section independence for limited dependent variable panel data models. *Oxford Bulletin of Economics and Statistics*, **74,** 253–277.

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.

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.

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.

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.

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.

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.

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

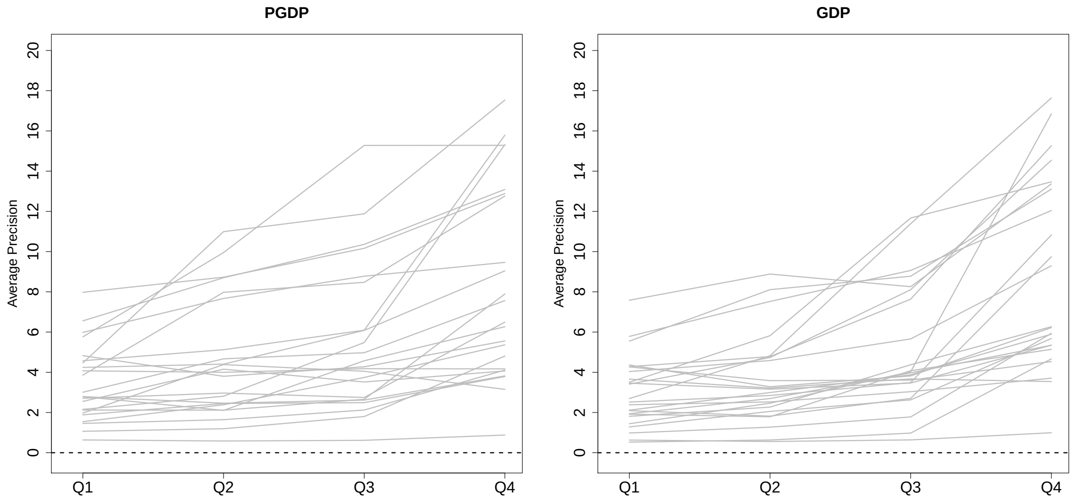
Sims, Christopher A (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50,** 665–690.

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

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.

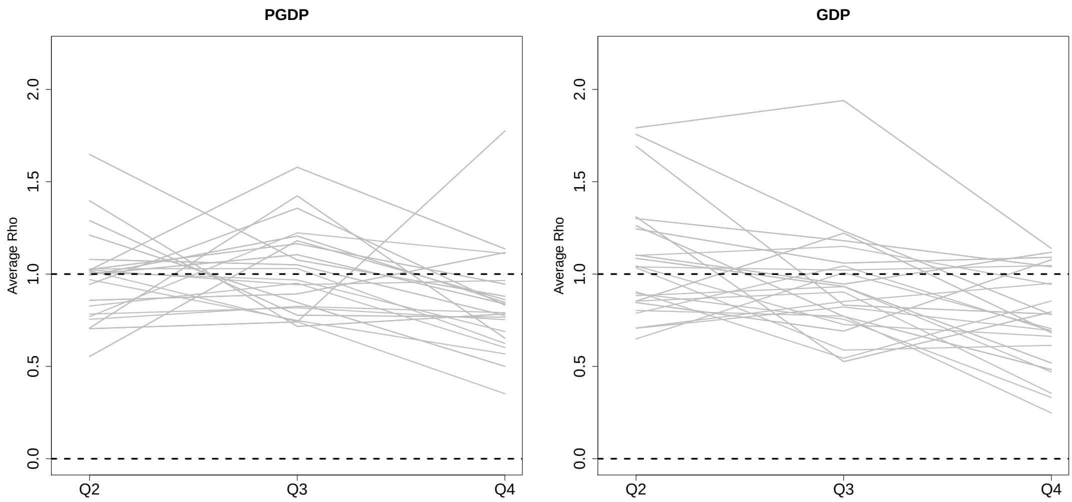
Zarnowitz, Victor and Lambros, Louis A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, **95,** 591–621.

Figure 1: **Precision of density forecasts**



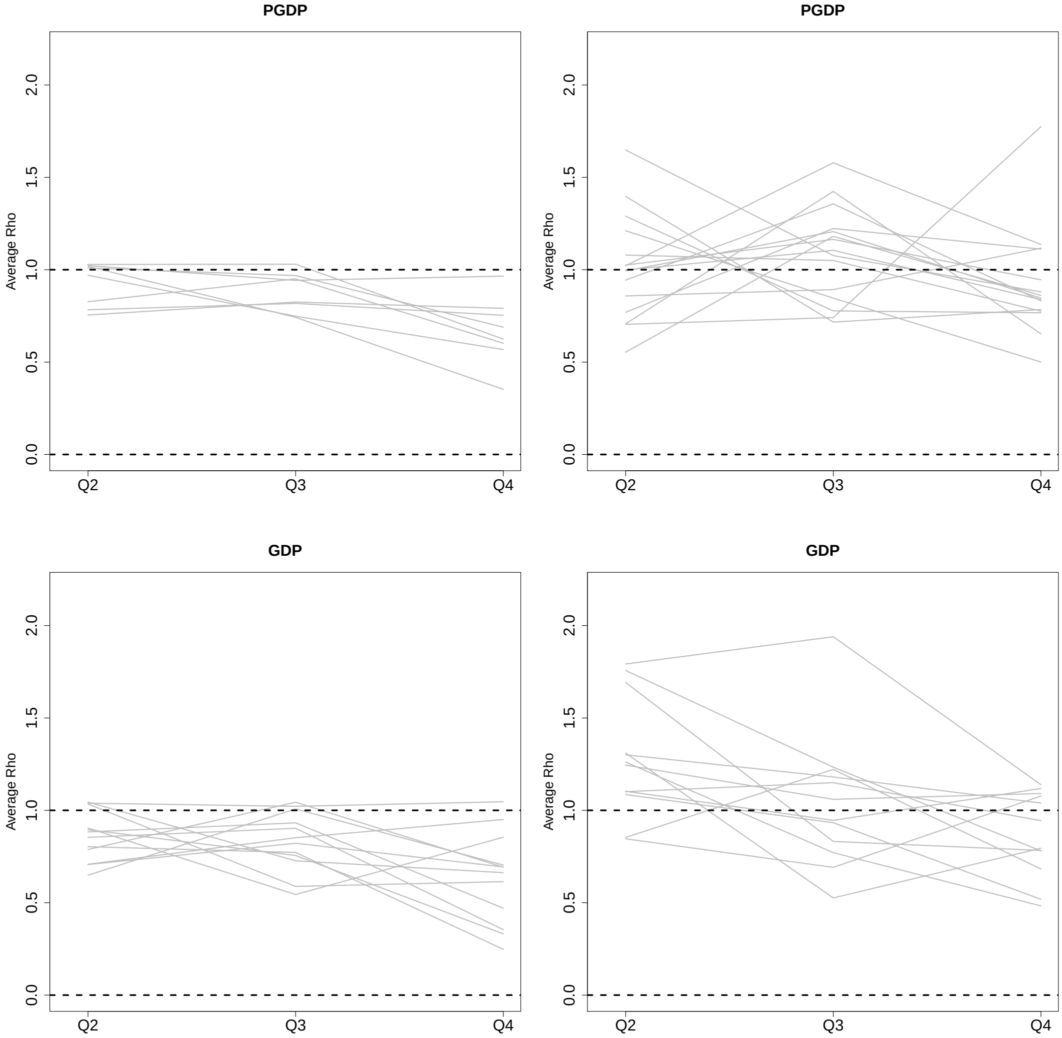
Each line represents the time-average precision (the inverse of the variance) of the density forecasts of a forecaster by quarter the forecast was issued. The left graph refers to forecast of the PGDP growth rate for the current year and includes 23 forecasters and the right graph refers to the GDP growth rate and includes 24 forecasters.

Figure 2: **Prior weight**



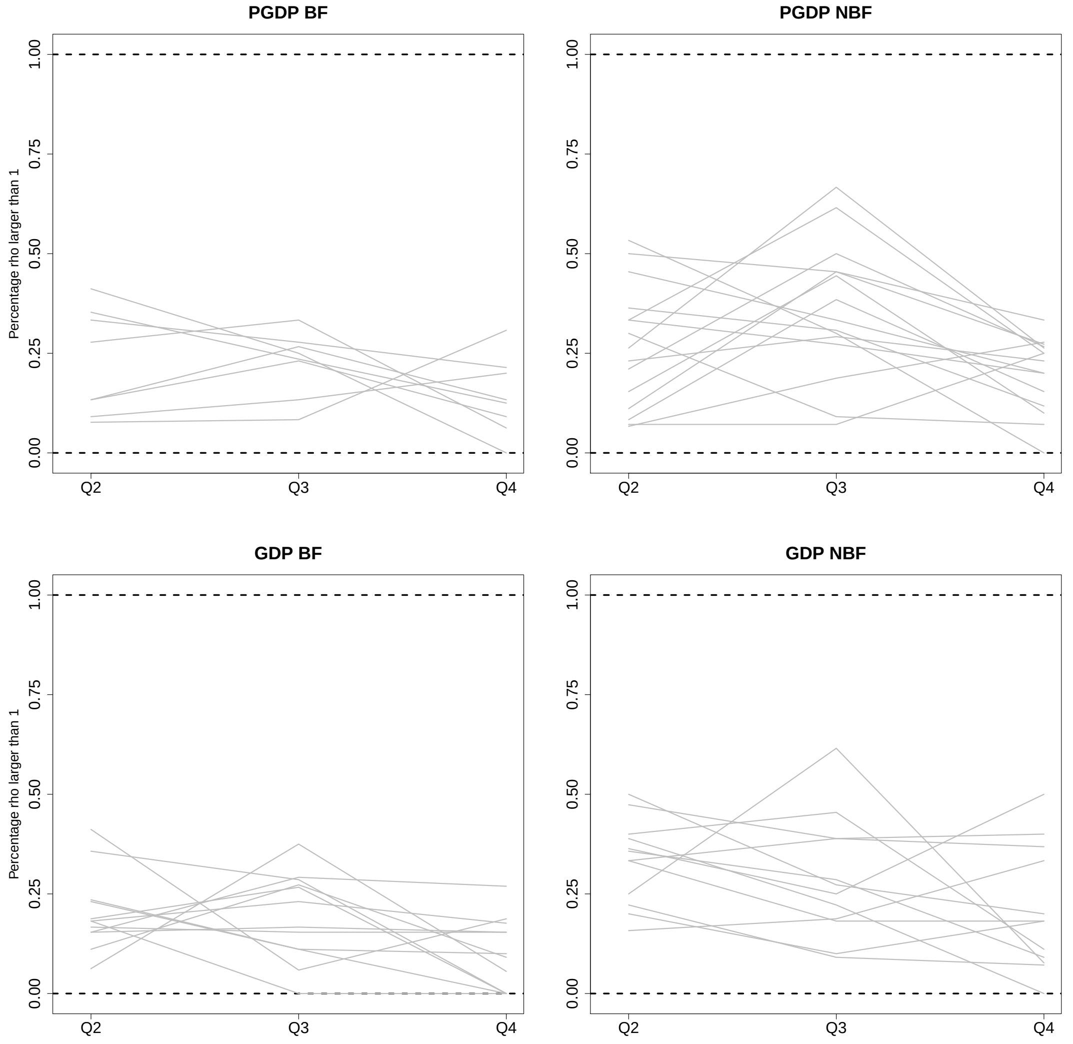
Each line represents the time-average prior weight of a forecaster by quarter with the weight defined as the ratio of the posterior to the prior precisions. The left graph refers to forecast of the PGDP growth rate for the current year and includes 23 forecasters and the right graph refers to the GDP growth rate and includes 24 forecasters.

Figure 3: **Prior weight (BF and NBF)**



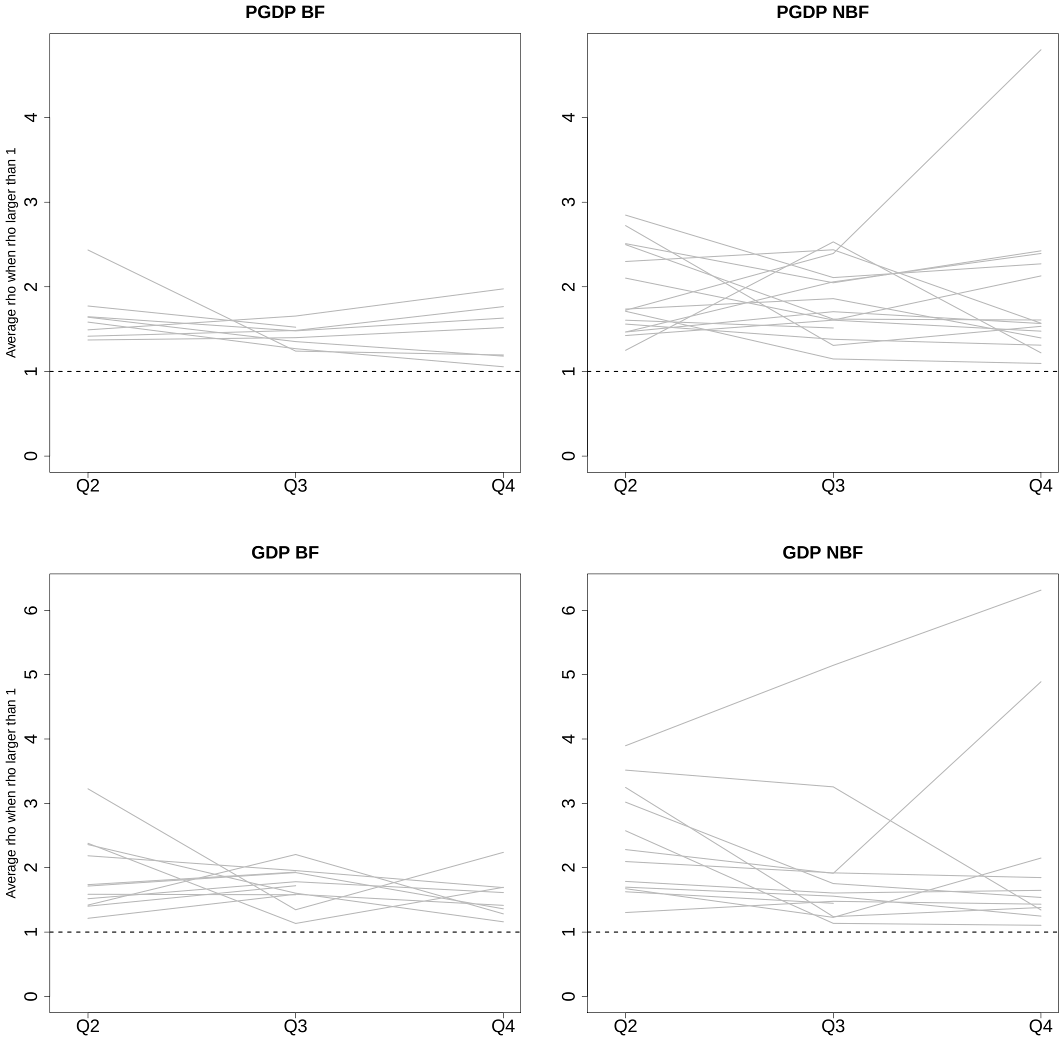
The graphs show the time-average prior weight by variable (PGDP, GDP) and also by type of forecaster (BF, NBF). We define as Bayesian Forecaster (BF) those forecasters that have prior weight $\bar{\rho}^i_{q,\cdot}$ lower than 1 at all horizons and as Non-Bayesian Forecasters (NBF) those that for at least one horizon have weight larger than 1. For PGDP we classify 8 forecasters as BF and 15 as NBF, while for GDP forecasters are equally split between BF and NBF.

Figure 4: **Fraction of prior weights larger than 1**



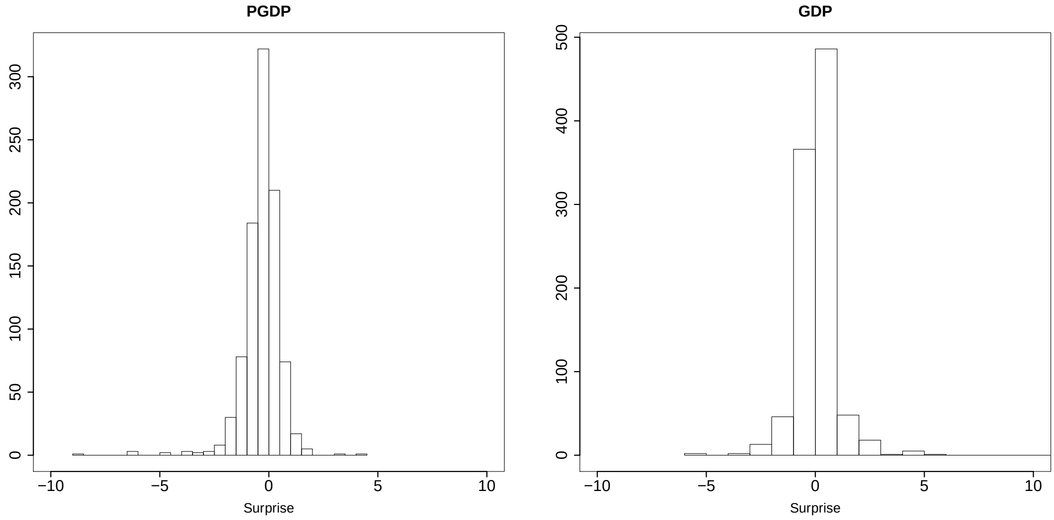
The graphs show, for each forecaster in our sample, the fraction of prior weights that are larger than 1 by variable (PGDP, GDP) and by forecaster type (BF, NBF).

Figure 5: **Average prior weight when weight larger than 1**



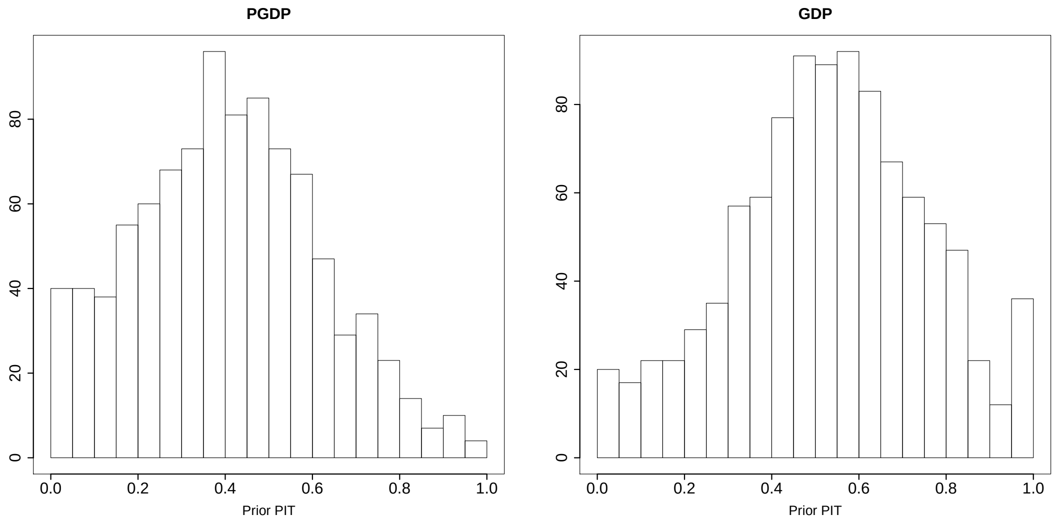
The graphs show the average weight for each forecaster when the weight is larger than 1 by variable (PGDP, GDP) and by forecaster type (BF, NBF).

Figure 6: **Distribution of surprises**



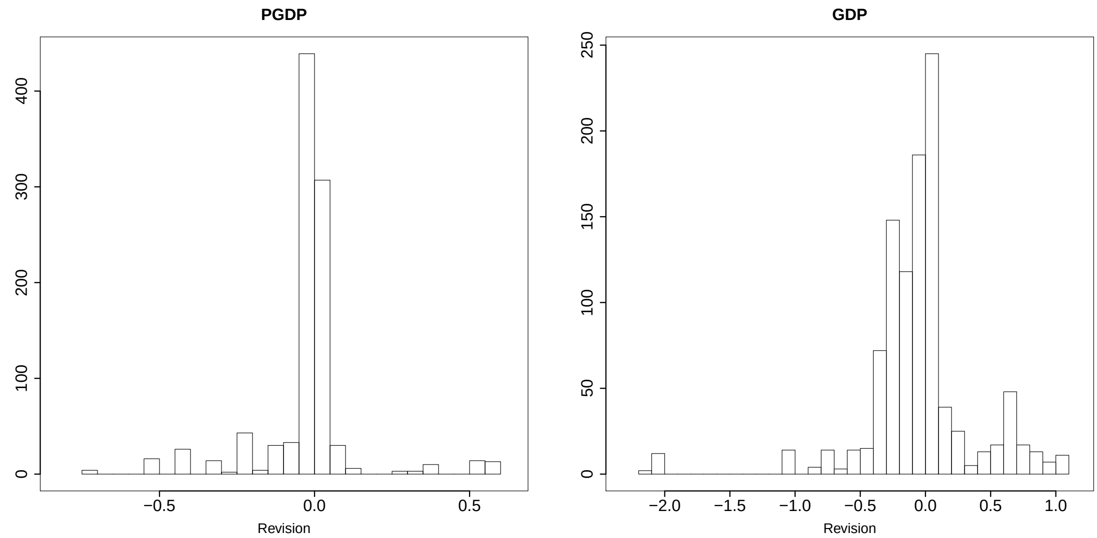
Histogram of the forecaster-specific surprise variable for PGDP (left) and GDP (right) defined in Equation (6). The histogram pools surprises across forecasters and time.

Figure 7: **Prior probability of surprises**



Histogram of the prior probability of the forecaster-specific unexpected part of the data release which is obtained by evaluating the forecaster's normal CDF in quarter q of year t at the surprise value. The histogram pools these probabilities across forecasters and time.

Figure 8: **Prior probability of revisions**



Histogram of the prior probability of the forecaster-specific news due to revisions to previously announced data which is obtained by evaluating the forecaster's normal CDF in quarter q of year t at the revised value. The histogram pools these probabilities across forecasters and time.

Table 1: **Weight regression (PGDP)**

Variable	Group	Model B	Model C	Model D	Model E	Model F	Model G
$\rho_{i,q-1,t}$	ALL	-0.14 (5)	-0.22 (7)	-0.23 (6)	-0.19 (6)	-0.24 (6)	-0.25 (5)
	BF	-0.14 (2)	-0.22 (2)	-0.25 (2)	-0.24 (2)	-0.26 (2)	-0.27 (2)
	NBF	-0.14 (3)	-0.22 (5)	-0.22 (4)	-0.17 (4)	-0.23 (4)	-0.23 (3)
$\psi_{i,q-1,t}$	ALL		0.11 (9)	0.1 (8)	0.09 (7)	0.09 (6)	0.09 (4)
	BF		0.1 (3)	0.09 (2)	0.09 (2)	0.1 (2)	0.1 (1)
	NBF		0.12 (6)	0.11 (6)	0.09 (5)	0.09 (4)	0.09 (3)
$S_{i,q,t}$	ALL			-0.09 (8)		-0.13 (5)	-0.12 (6)
	BF			0.05 (2)		0.01 (1)	0.01 (1)
	NBF			-0.17 (6)		-0.2 (4)	-0.2 (5)
$S^{+}_{i,q,t}$	ALL				0.21 (4)		
	BF				0.15 (1)		
	NBF				0.23 (3)		
$S^{-}_{i,q,t}$	ALL				-0.27 (10)		
	BF				-0.05 (2)		
	NBF				-0.38 (8)		
$I^{lowS}_{i,q,t}$	ALL					0.16 (5)	0.18 (5)
	BF					-0.02 (1)	-0.05 (1)
	NBF					0.25 (4)	0.3 (4)
$I^{highS}_{i,q,t}$	ALL					0.53 (7)	0.48 (7)
	BF					-0.03 (2)	-0.06 (2)
	NBF					0.82 (5)	0.77 (5)
$I^{lowR}_{i,q,t}$	ALL						0.1 (2)
	BF						0.15 (1)
	NBF						0.07 (1)
$I^{highR}_{i,q,t}$	ALL						-0.04 (1)
	BF						0.03 (0)
	NBF						-0.07 (1)
	Model A						
CD stat	1.104	0.994	1.709	1.312	0.711	0.698	-0.27
Av. Corr.	0.006	0.003	0.017	0.002	0.001	0.018	-0.014
				R^2			
ALL average	0.73	0.74	0.77	0.79	0.81	0.82	0.83
ALL min/max	0.5/0.92	0.51/0.92	0.51/0.92	0.56/0.92	0.56/0.94	0.56/0.95	0.56/0.96
BF average	0.8	0.81	0.83	0.84	0.85	0.85	0.86
NBF average	0.7	0.71	0.73	0.76	0.79	0.81	0.81
				AIC/BIC			
AIC ALL	2/8	0/1	2/2	8/4	2/3	7/4	2/1
AIC BF	1/3	0/1	1/1	4/1	1/1	1/1	0/0
AIC NBF	1/5	0/0	1/1	4/3	1/2	6/3	2/1

Regression results for Equation (5) for several specifications denoted by Model A to G. In Model A we only include the quarterly dummy variables while for Models B to G we include additional variables as reported in each column of the Table. We estimate the models by OLS for each forecaster and report the Mean Group (MG) estimate as well as the number of forecasters for which the coefficient is significant at 10% level. The row ALL denotes that the MG estimate is taken across the 23 forecasters in the sample while for BF and NBF the mean is taken over the 8 and 15 forecasters classified in each group. *Av. Corr.* = average cross-correlation of the forecaster-specific residuals, *CD stat* = Cross-Section Dependence test in Equation (9) which is standard normally distributed. We also report the average/min/max of the R^2 across forecasters for ALL, BF, and NBF and the bottom three rows provide the number of forecasters for which each model is selected according to AIC and BIC.

Table 2: **Weight regression (GDP)**

Variable	Group	Model B	Model C	Model D	Model E	Model F	Model G
$\rho_{i,q-1,t}$	ALL	-0.18 (8)	-0.32 (13)	-0.21 (10)	-0.2 (8)	-0.24 (9)	-0.28 (7)
	BF	-0.16 (4)	-0.28 (5)	-0.18 (5)	-0.16 (3)	-0.16 (3)	-0.2 (3)
	NBF	-0.2 (4)	-0.37 (8)	-0.24 (5)	-0.24 (5)	-0.33 (6)	-0.35 (4)
$\psi_{i,q-1,t}$	ALL		0.17 (15)	0.03 (5)	0.02 (6)	0.05 (5)	0.04 (5)
	BF		0.12 (6)	0.03 (3)	0.01 (3)	0 (1)	-0.01 (1)
	NBF		0.22 (9)	0.04 (2)	0.04 (3)	0.11 (4)	0.1 (4)
$S_{i,q,t}$	ALL			-0.08 (5)		-0.02 (5)	-0.04 (7)
	BF			0 (2)		0.1 (3)	0.12 (2)
	NBF			-0.16 (3)		-0.14 (2)	-0.21 (5)
$S^+_{i,q,t}$	ALL				0.02 (4)		
	BF				0.2 (3)		
	NBF				-0.16 (1)		
$S^-_{i,q,t}$	ALL				-0.25 (4)		
	BF				-0.23 (2)		
	NBF				-0.26 (2)		
$I^{lowS}_{i,q,t}$	ALL					0.13 (5)	0.16 (8)
	BF					0.4 (2)	0.64 (3)
	NBF					-0.15 (3)	-0.32 (5)
$I^{highS}_{i,q,t}$	ALL					-0.26 (6)	-0.11 (1)
	BF					-0.2 (4)	-0.37 (0)
	NBF					-0.33 (2)	0.14 (1)
$I^{lowR}_{i,q,t}$	ALL						-0.21 (4)
	BF						-0.73 (2)
	NBF						0.3 (2)
$I^{highR}_{i,q,t}$	ALL						-0.36 (2)
	BF						-0.07 (1)
	NBF						-0.66 (1)
	Model A						
CD stat	4.684	-2.484	-1.846	-0.949	-1.962	-1.599	-1.785
Av. Corr.	0.055	-0.027	-0.021	-0.019	-0.034	-0.028	-0.03
				R^2			
ALL average	0.66	0.74	0.78	0.85	0.87	0.9	0.93
ALL min/max	0.25/0.88	0.5/0.89	0.54/0.91	0.68/0.95	0.7/0.96	0.71/0.97	0.72/1
BF average	0.7	0.77	0.8	0.86	0.88	0.89	0.93
NBF average	0.62	0.7	0.76	0.84	0.86	0.9	0.93
				AIC/BIC			
ALL	0/1	0/0	0/1	5/11	5/4	1/0	13/7
BF	0/1	0/0	0/1	2/5	1/1	1/0	8/4
NBF	0/0	0/0	0/0	3/6	4/3	0/0	5/3

Regression results for Equation (5) for several specifications denoted by Model A to G. In Model A we only include the quarterly dummy variables while for Models B to G we include additional variables as reported in each column of the Table. We estimate the models by CCE for each forecaster (except for Model A for which we use OLS) and report the Mean Group (MG) estimate as well as the number of forecasters for which the coefficient is significant at 10% level. The row ALL denotes that the MG estimate is taken across the 24 forecasters in the sample while for BF and NBF the mean is taken over the 12 forecasters classified in each group. *Av. Corr.* = average cross-correlation of the forecaster-specific residuals, *CD stat* = Cross-Section Dependence test in Equation (9) which is standard normally distributed. We also report the average/min/max of the R^2 across forecasters for ALL, BF, and NBF and the bottom three rows provide the number of forecasters for which each model is selected according to AIC and BIC.

Table 3: **Probit regression (PGDP)**

Variable	Group	Model A	Model B	Model C	Model D	Model E	Model F	Model G
$\rho_{i,q-1,t}$	ALL		-0.62 (11)	-1.31 (17)	-1.45 (17)	-2.79 (13)	-1.82 (14)	-1.93 (15)
	BF		-0.64 (4)	-1.34 (7)	-1.24 (6)	-3.75 (4)	-1.65 (5)	-1.55 (6)
	NBF		-0.61 (7)	-1.29 (10)	-1.56 (11)	-2.28 (9)	-1.91 (9)	-2.14 (9)
$\psi_{i,q-1,t}$	ALL			-0.51 (3)	-0.63 (4)	-2.32 (3)	-0.63 (1)	-0.86 (1)
	BF			-0.67 (2)	-0.56 (1)	-4.65 (2)	0.01 (0)	0.63 (0)
	NBF			-0.43 (1)	-0.67 (3)	-1.08 (1)	-0.97 (1)	-1.65 (1)
$S_{i,q,t}$	ALL				-1.33 (8)		-1.88 (9)	-2.06 (7)
	BF				-1.38 (2)		-1.95 (2)	-1.79 (2)
	NBF				-1.31 (6)		-1.84 (7)	-2.21 (5)
$S^{+}_{i,q,t}$	ALL					-2.09 (9)		
	BF					-2.41 (4)		
	NBF					-1.91 (5)		
$S^{-}_{i,q,t}$	ALL					-1.01 (3)		
	BF					-1.82 (1)		
	NBF					-0.58 (2)		
$I^{lowS}_{i,q,t}$	ALL						-1.16 (4)	-1.7 (6)
	BF						-2.4 (3)	-3.96 (3)
	NBF						-0.49 (1)	-0.49 (3)
$I^{highS}_{i,q,t}$	ALL						0.37 (7)	0.36 (8)
	BF						0.42 (3)	0.39 (4)
	NBF						0.34 (4)	0.34 (4)
$I^{lowR}_{i,q,t}$	ALL							-0.4 (1)
	BF							-0.02 (1)
	NBF							-0.61 (0)
$I^{highR}_{i,q,t}$	ALL							0.17 (0)
	BF							-0.32 (0)
	NBF							0.43 (0)
		Model A						
CD stat		1.496	1.641	1.91	1.706	1.294	1.93	1.373
Av. Corr.		0.014	0.009	0.02	0.011	0.005	0.021	0.008
				Fraction of correct predictions				
ALL average		0.62	0.63	0.66	0.68	0.72	0.74	0.77
ALL min/max		0.46/0.77	0.51/0.79	0.5/0.79	0.52/0.83	0.52/0.93	0.52/0.89	0.59/1
BF average		0.58	0.61	0.64	0.67	0.72	0.76	0.81
NBF average		0.64	0.65	0.67	0.69	0.73	0.72	0.75
					AIC/BIC			
ALL		7/14	0/0	2/2	1/4	4/2	3/0	6/1
BF		1/5	0/0	1/0	1/1	1/1	0/0	4/1
NBF		6/9	0/0	1/2	0/3	3/1	3/0	2/0

Results for the probit regression with dependent variable $I(\rho^{i}_{q,t} > 1)$ and using the same specifications as in Tables (1) and (2). We estimate the models by ML for each forecaster and report the mean estimate across forecasters as well as the number of forecasters for which the coefficient is significant at 10% level. The row ALL denotes that the mean estimate is taken across the 23 forecasters in the sample while for BF and NBF the mean is taken over the 8 and 15 forecasters classified in each group. *Av. Corr.* = average cross-correlation of the forecaster-specific residuals, *CD stat* = Cross-Section Dependence test in Equation (9) which is standard normally distributed. We also report the average/min/max of the fraction of correct predictions across forecasters for ALL, BF, and NBF and the bottom three rows provide the number of forecasters for which each model is selected according to AIC and BIC.

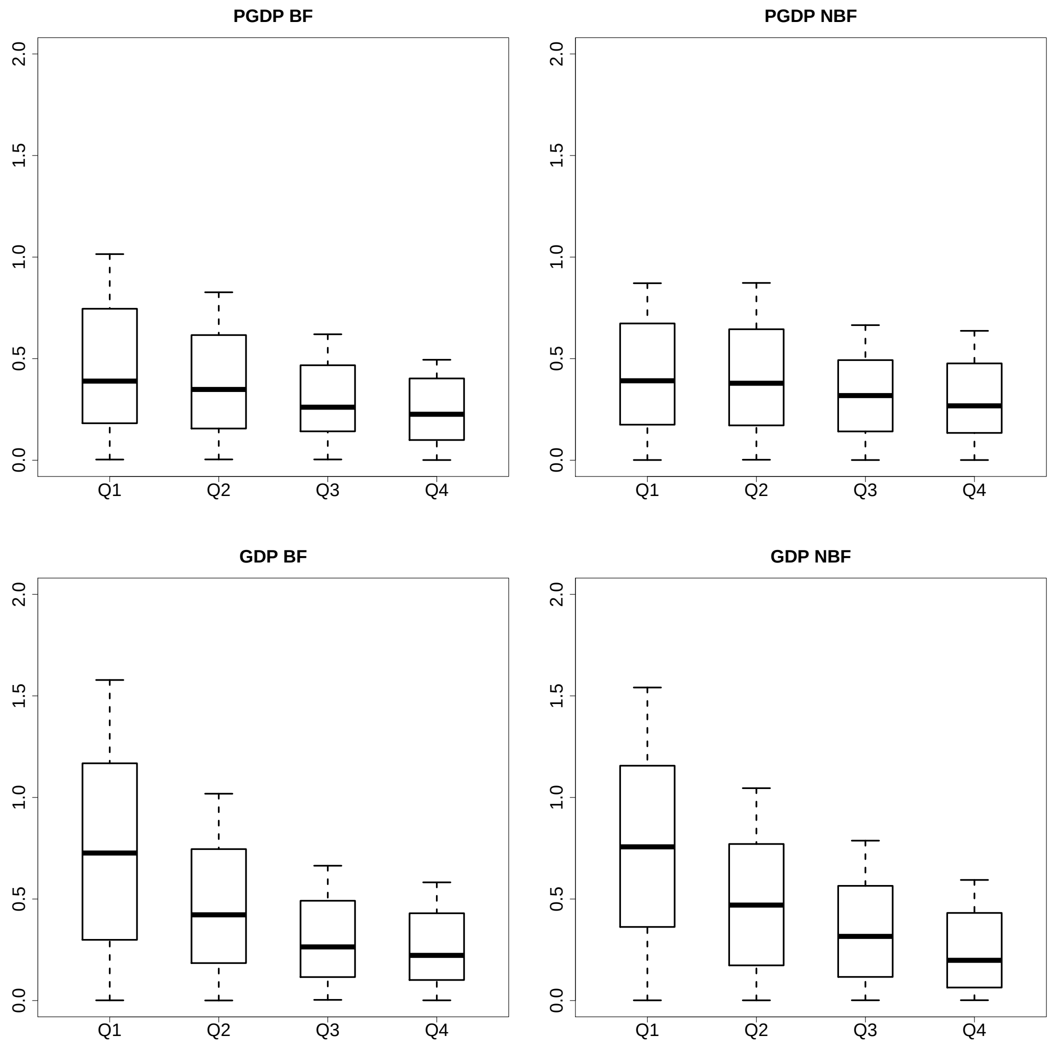
Table 4: **Probit regression (GDP)**

Variable	Group	Model B	Model C	Model D	Model E	Model F	Model G
$\rho_{i,q-1,t}$	ALL	-0.49 (10)	-0.99 (13)	-1.05 (14)	-1.38 (17)	-1.19 (14)	-1.22 (13)
	BF	-0.67 (7)	-1.2 (8)	-1.3 (9)	-1.7 (8)	-1.53 (8)	-1.54 (8)
	NBF	-0.31 (3)	-0.79 (5)	-0.79 (5)	-1.05 (9)	-0.85 (6)	-0.9 (5)
$\psi_{i,q-1,t}$	ALL		-0.54 (2)	-0.43 (3)	-0.77 (4)	-0.4 (4)	-0.84 (3)
	BF		-0.89 (1)	-0.61 (2)	-0.86 (2)	-0.49 (2)	-0.72 (2)
	NBF		-0.2 (1)	-0.25 (1)	-0.67 (2)	-0.32 (2)	-0.96 (1)
$S_{i,q,t}$	ALL			-1.73 (5)		-2.2 (6)	-2.71 (6)
	BF			-2.54 (3)		-3.5 (4)	-3.62 (4)
	NBF			-0.91 (2)		-0.91 (2)	-1.8 (2)
$S^+_{i,q,t}$	ALL				-2.5 (8)		
	BF				-3.81 (5)		
	NBF				-1.19 (3)		
$S^-_{i,q,t}$	ALL				-1.78 (8)		
	BF				-2.5 (3)		
	NBF				-1.07 (5)		
$I^{lowS}_{i,q,t}$	ALL					-1.72 (9)	-1.22 (9)
	BF					-2.4 (4)	-2.14 (4)
	NBF					-1.03 (5)	-0.29 (5)
$I^{highS}_{i,q,t}$	ALL					0.29 (6)	0.31 (5)
	BF					0.32 (2)	0.32 (2)
	NBF					0.27 (4)	0.29 (3)
$I^{lowR}_{i,q,t}$	ALL						0.19 (4)
	BF						0.34 (1)
	NBF						0.23 (3)
$I^{highR}_{i,q,t}$	ALL						-0.24 (1)
	BF						-0.89 (0)
	NBF						0.41 (1)

	Model A						
CD stat	4.586	4.804	5.093	4.563	4.41	4.836	3.766
Av. Corr.	0.062	0.067	0.065	0.06	0.057	0.067	0.053
			Fraction of correct predictions				
ALL average	0.59	0.63	0.69	0.69	0.72	0.74	0.76
ALL min/max	0.44/0.78	0.45/0.78	0.56/0.78	0.52/0.85	0.6/1	0.6/1	0.66/1
BF average	0.59	0.61	0.68	0.68	0.73	0.75	0.78
NBF average	0.58	0.64	0.69	0.7	0.7	0.72	0.73
				AIC/BIC			
ALL	7/18	0/3	2/2	8/0	4/0	2/1	1/0
AIC BF	3/8	0/1	1/2	4/0	1/0	2/1	1/0
AIC NBF	410	0/2	1/0	4/0	3/0	0/0	0/0

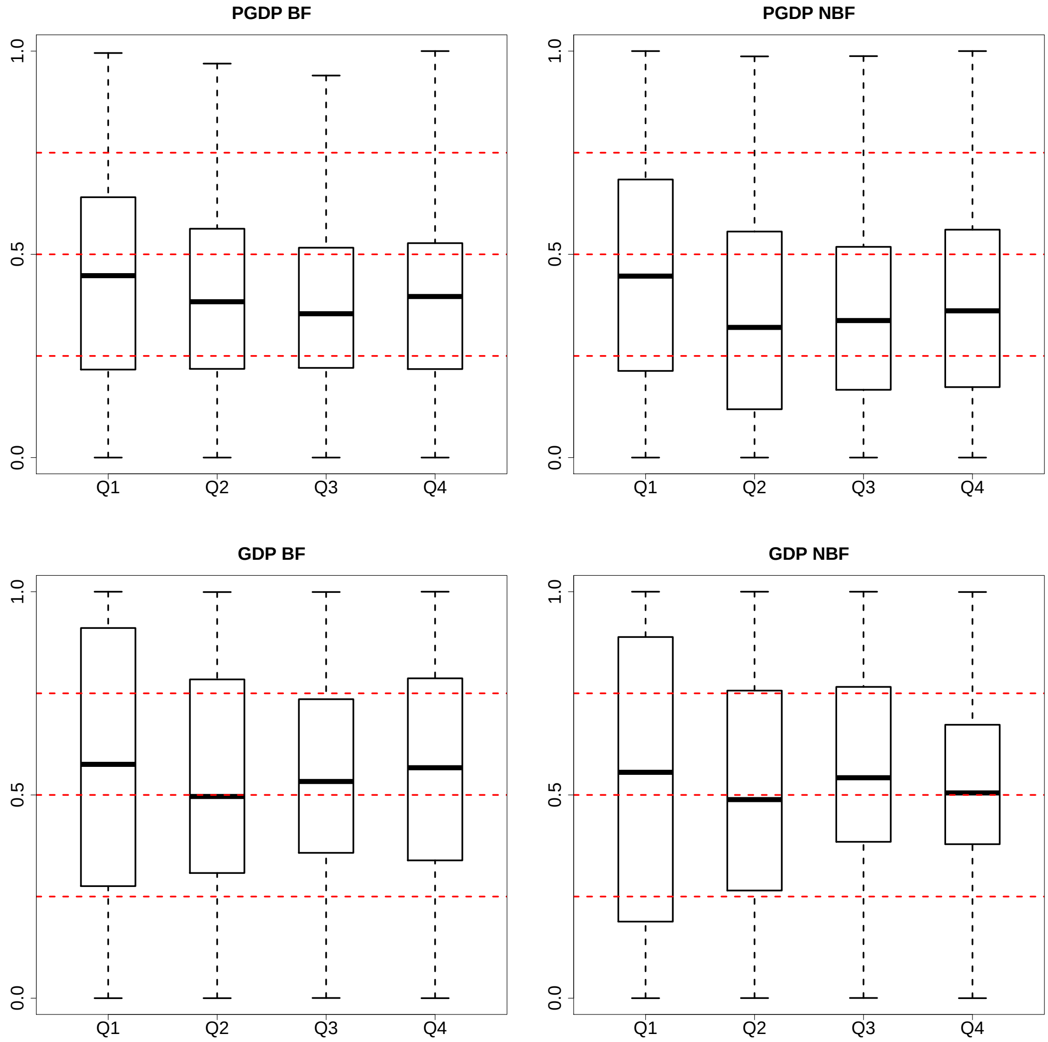
Results for the probit regression with dependent variable $I(\rho^i_{q,t} > 1)$ and using the same specifications as in Tables (1) and (2). We estimate the models by ML for each forecaster and report the mean estimate across forecasters as well as the number of forecasters for which the coefficient is significant at 10% level. The row ALL denotes that the mean estimate is taken across the 24 forecasters in the sample while for BF and NBF the mean is taken over the 12 forecasters classified in each group. *Av. Corr.* = average cross-correlation of the forecaster-specific residuals, *CD stat* = Cross-Section Dependence test in Equation (9) which is standard normally distributed. We also report the average/min/max of the fraction of correct predictions across forecasters for ALL, BF, and NBF and the bottom three rows provide the number of forecasters for which each model is selected according to AIC and BIC.

Figure 9: **Distribution of the absolute forecast errors**



The graphs show boxplots of the absolute forecast error by variable (PGDP, GDP) and by forecaster type (BF, NBF) obtained by pooling the forecast errors across forecasters and time. The inner box represents the first, second and third quartile of the distribution and the whiskers represent the minimum and maximum.

Figure 10: **Distribution of PIT**



The graphs show boxplots of the prior probability of the realization by variable (PGDP, GDP) and by forecaster type (BF, NBF) obtained by pooling the forecast errors across forecasters and time. The probability is calculated by evaluating the forecaster's normal CDF at the realized value. The inner box represents the first, second and third quartile of the distribution and the whiskers represent the minimum and maximum. The horizontal dashed lines represent the 0.25, 0.5, and 0.75 probability levels.

Table 5: **Absolute Forecast Error regression (PGDP)**

Variable	Group	Model B	Model C	Model D	Model E	Model F	Model G
$\lvert e_{i,q-1,t}\rvert$	ALL	0.22 (11)	0.19 (9)	0.26 (9)	0.29 (8)	0.27 (8)	0.27 (8)
	BF	0.17 (2)	0.14 (2)	0.19 (4)	0.22 (4)	0.23 (3)	0.23 (3)
	NBF	0.25 (9)	0.22 (7)	0.3 (5)	0.33 (4)	0.29 (5)	0.29 (5)
$\psi_{i,q-1,t}$	ALL		0 (0)	0.02 (2)	0.01 (3)	0.01 (3)	-0.01 (2)
	BF		0.01 (0)	0.02 (1)	0.03 (2)	0.03 (2)	0.02 (1)
	NBF		-0.01 (0)	0.01 (1)	0 (1)	0 (1)	-0.02 (1)
$S_{i,q,t}$	ALL			0 (2)		-0.04 (1)	-0.06 (0)
	BF			0.02 (2)		-0.02 (1)	-0.01 (0)
	NBF			-0.01 (0)		-0.06 (0)	-0.08 (0)
$S^{+}_{i,q,t}$	ALL				-0.11 (3)		
	BF				-0.13 (1)		
	NBF				-0.11 (2)		
$S^{-}_{i,q,t}$	ALL				0.03 (4)		
	BF				0.04 (1)		
	NBF				0.03 (3)		
$I^{lowS}_{i,q,t}$	ALL					-0.12 (4)	-0.18 (3)
	BF					-0.22 (0)	-0.2 (0)
	NBF					-0.07 (4)	-0.18 (3)
$I^{highS}_{i,q,t}$	ALL					0.03 (1)	-0.05 (1)
	BF					0 (0)	0.11 (0)
	NBF					0.05 (1)	-0.14 (1)
$I^{lowR}_{i,q,t}$	ALL						-0.05 (2)
	BF						-0.09 (0)
	NBF						-0.02 (2)
$I^{highR}_{i,q,t}$	ALL						0.02 (1)
	BF						-0.1 (1)
	NBF						0.08 (0)
	Model A						
CD stat	13.239	-4.019	-4.03	-2.76	-2.11	-1.804	-2.101
Av. Corr.	0.123	-0.043	-0.045	-0.046	-0.042	-0.032	-0.034
				R^2			
ALL average	0.6	0.73	0.75	0.79	0.81	0.82	0.85
ALL min/max	0.44/0.74	0.54/0.85	0.56/0.86	0.55/0.93	0.61/0.94	0.59/0.96	0.66/0.96
BF average	0.57	0.7	0.72	0.76	0.78	0.79	0.82
NBF average	0.62	0.75	0.76	0.81	0.83	0.84	0.86
				AIC/BIC			
AIC ALL	0/0	1/5	0/0	10/14	6/3	4/1	2/0
AIC BF	0/0	0/1	0/0	5/6	2/1	0/0	1/0
AIC NBF	0/0	1/4	0/0	5/8	4/2	4/1	1/0

Regression results for Equation (10) for several specifications denoted by Model A to G. In Model A we only include the quarterly dummy variables while for Models B to G we include additional variables as reported in each column of the Table. We estimate the models by CCE for each forecaster and report the Mean Group (MG) estimate as well as the number of forecasters for which the coefficient is significant at 10% level. The row ALL denotes that the MG estimate is taken across the 23 forecasters in the sample while for BF and NBF the mean is taken over the 8 and 15 forecasters classified in each group. *Av. Corr.* = average cross-correlation of the forecaster-specific residuals, *CD stat* = Cross-Section Dependence test in Equation (9) which is standard normally distributed. We also report the average/min/max of the R^2 across forecasters for ALL, BF, and NBF and the bottom three rows provide the number of forecasters for which each model is selected according to AIC and BIC.

Table 6: **Absolute Forecast Error regression (GDP)**

Variable	Group	Model B	Model C	Model D	Model E	Model F	Model G
$\|e_{i,q-1,t}\|$	ALL	0.18 (9)	0.16 (8)	0.19 (11)	0.23 (11)	0.21 (10)	0.2 (10)
	BF	0.16 (3)	0.15 (3)	0.18 (6)	0.22 (5)	0.19 (4)	0.2 (4)
	NBF	0.19 (6)	0.17 (5)	0.21 (5)	0.24 (6)	0.22 (6)	0.2 (6)
$\psi_{i,q-1,t}$	AL		0.02 (0)	0.02 (1)	0.03 (3)	0.01 (3)	0.02 (3)
	BF		0.01 (0)	0 (1)	0 (2)	-0.02 (2)	-0.01 (1)
	NBF		0.03 (0)	0.05 (0)	0.05 (1)	0.05 (1)	0.06 (2)
$S_{i,q,t}$	ALL			0 (5)		-0.05 (3)	-0.04 (2)
	BF			0.02 (3)		-0.03 (2)	-0.01 (1)
	NBF			-0.02 (2)		-0.06 (1)	-0.07 (1)
$S^{+}_{i,q,t}$	ALL				-0.04 (1)		
	BF				0 (0)		
	NBF				-0.08 (1)		
$S^{-}_{i,q,t}$	ALL				0.04 (3)		
	BF				0.04 (0)		
	NBF				0.04 (3)		
$I^{lowS}_{i,q,t}$	ALL					-0.12 (3)	-0.06 (4)
	BF					-0.06 (0)	0 (0)
	NBF					-0.17 (3)	-0.12 (4)
$I^{highS}_{i,q,t}$	ALL					0.15 (4)	0.06 (3)
	BF					0.16 (1)	0.04 (1)
	NBF					0.13 (3)	0.08 (2)
$I^{lowR}_{i,q,t}$	ALL						-0.01 (2)
	BF						0.08 (1)
	NBF						-0.11 (1)
$I^{highR}_{i,q,t}$	ALL						-0.31 (3)
	BF						-0.58 (1)
	NBF						-0.04 (2)
	Model A						
CD stat	38.936	-3.517	-3.796	-2.565	-2.289	-1.278	-0.637
Av. Corr.	0.367	-0.042	-0.051	-0.039	-0.034	-0.02	-0.009
				R^2			
ALL average	0.61	0.83	0.83	0.84	0.85	0.87	0.89
ALL min/max	0.51/0.83	0.66/0.91	0.67/0.91	0.76/0.92	0.78/0.93	0.8/0.94	0.81/0.96
BF average	0.62	0.83	0.83	0.85	0.86	0.88	0.89
NBF average	0.61	0.83	0.83	0.82	0.84	0.86	0.88
				AIC/BIC			
ALL	0/0	2/5	0/0	9/16	4/0	5/0	4/3
BF	0/0	1/2	0/0	6/9	2/0	2/0	1/1
NBF	0/0	1/3	0/0	3/7	2/0	3/0	3/2

Regression results for Equation (10) for several specifications denoted by Model A to G. In Model A we only include the quarterly dummy variables while for Models B to G we include additional variables as reported in each column of the Table. We estimate the models by CCE for each forecaster and report the Mean Group (MG) estimate as well as the number of forecasters for which the coefficient is significant at 10% level. The row ALL denotes that the MG estimate is taken across the 24 forecasters in the sample while for BF and NBF the mean is taken over the 12 forecasters classified in each group. *Av. Corr.* = average cross-correlation of the forecaster-specific residuals, *CD stat* = Cross-Section Dependence test in Equation (9) which is standard normally distributed. We also report the average/min/max of the R^2 across forecasters for ALL, BF, and NBF and the bottom three rows provide the number of forecasters for which each model is selected according to AIC and BIC.