



EUROPEAN CENTRAL BANK

EUROSYSTEM



In 2014 all ECB publications feature a motif taken from the €20 banknote.



WORKING PAPER SERIES

NO 1679 / MAY 2014

DENSITY CHARACTERISTICS AND DENSITY FORECAST PERFORMANCE

A PANEL ANALYSIS

Geoff Kenny, Thomas Kostka
and Federico Masera

NOTE: This Working Paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Geoff Kenny (corresponding author)

European Central Bank; e-mail: geoff.kenny@ecb.europa.eu

Thomas Kostka

European Central Bank

Federico Masera

Universidad Carlos III de Madrid

© European Central Bank, 2014

Address	Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
Postal address	Postfach 16 03 19, 60066 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Internet	http://www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors. This paper can be downloaded without charge from <http://www.ecb.europa.eu> or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=2436416. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, <http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html>

ISSN	1725-2806 (online)
ISBN	978-92-899-1087-3
EU Catalogue No	QB-AR-14-053-EN-N (online)

ABSTRACT

In this paper, we exploit micro data from the ECB Survey of Professional Forecasters (SPF) to examine the link between the characteristics of macroeconomic density forecasts (such as their location, spread, skewness and tail risk) and density forecast performance. Controlling for the effects of common macroeconomic shocks, we apply cross-sectional and fixed effect panel regressions linking such density characteristics and density forecast performance. Our empirical results suggest that many macroeconomic experts could systematically improve their density performance by correcting a downward bias in their variances. Aside from this shortcoming in second moment characteristics of the individual densities, other higher moment features, such as skewness or variation in the degree of probability mass given to the tails of the predictive distributions tend - as a rule - not to contribute significantly to enhancing individual density forecast performance.

Keywords: Density forecasting, Forecast evaluation, Survey of Professional Forecasters, Panel data

JEL: C22, C53

NON-TECHNICAL SUMMARY

In this paper we exploit the microeconomic information contained in the European Central Bank (ECB) Survey of Professional Forecasters (SPF) as a means to contribute further to understanding the predictive performance of surveyed density forecasts. In particular, we examine the possible role of density features such as their location, spread, skewness and tail risk in determining density forecast performance both over time and across individuals. Understanding this link is of relevance to both forecast users – such as the ECB - and forecast producers. In particular, the insights from our study are of interest to survey users who may rely on specific density features when evaluating different policy choices (e.g. relating to tail risks or possible skewness in the distributions). Moreover, density forecast producers, including those forecast producers responding to the SPF questionnaire, can potentially improve their density forecast performance by gaining an understanding about how their density features have impacted on their historical density performance.

Our analysis focuses on the one and two-year horizon density forecasts for euro area real output growth and consumer price inflation. We begin by constructing individual measures of density forecast performance from this dataset. Our preferred performance measure is the Ranked Probability Score (RPS) which is based on the entire predictive distribution and rewards forecasters who concentrate a high probability mass in regions of their density that are close to where the outcome occurs. Next, we estimate directly at the individual level key characteristics of the SPF densities such as their means and higher moment features such as their variances, their skewness and tail probability mass. We then propose a set of cross-sectional and fixed effect panel regressions to examine the role of key distributional features in explaining density forecast performance both across time and across individuals. Controlling for the role of differences in point forecast accuracy (density location) as well as for other common shocks impacting on aggregate density performance, these regressions help shed light on whether or not higher order density characteristics, such as variance, skewness or the fatness of a density forecast's tails can contribute to improving forecast performance. Such a mode of analysis, responds to a clear need to generate empirical evidence concerning the quality of information that is contained in such density forecast features. Importantly, our analysis also sheds light on the use of judgement in forecasting and its impact on the quality of macroeconomic forecasts. In particular, in a recent questionnaire sent to the participants in the ECB SPF, a large majority of respondents (over 80%) indicated that their reported probability distributions are derived either purely on the basis of judgement or from models with judgemental adjustments.

Our main findings highlight the importance of mean forecast accuracy as a systematic determinant of individual density forecast performance: Experts with less accurate mean forecasts in general also have less accurate density forecasts. However, our empirical results suggest that many macroeconomic experts are operating in a way in which density performance could be systematically improved by correcting a downward bias in their variances. In other words, many experts responding to the SPF are underestimating uncertainty and could improve their density performance by simply increasing their variances. Aside from this shortcoming in the spread of the individual expert densities, other features, such as density skewness or kurtosis generally often are found not to contribute significantly to enhancing individual density performance. Hence our results cast some doubt on the degree to which central banks should rely on these features of private sector density forecasts when evaluating macroeconomic risks. Increases in distributional skewness in absolute terms, for example, are invariably associated with a worse density performance – an effect which is statistically significant in our panel regressions for both GDP growth and consumer price inflation. As regards the assessment of tail risks, for the case of the relatively longer horizon density forecasts for GDP growth, we observe that a higher probability mass in the tails of the distributions is associated with a better density performance.

Overall our analysis points to some scope for professional forecasters to potentially improve their density forecast performance, in particular by correcting upwards a noticeable downward bias in their assessments of uncertainty. . Moreover our results suggest that forecast users should exercise due caution when extracting information from higher moments of the SPF densities such as measures of skewness and tail risk. Looking forward, with the expansion of the SPF dataset over time, it would be desirable to apply other density evaluation techniques and to possibly extend the analysis conducted here to consider other density scoring rules.

1. INTRODUCTION

The importance of reliable quantitative information about forecast densities - and not just point predictions - is widely recognized in both the econometric and applied forecasting literatures.¹ At the same time, recent applied research has revealed considerable shortcomings in the density forecasts of macroeconomists as collected in surveys. For example, Giordani and Söderlind (2006) explore the possible role of surveyed densities in explaining the equity premium puzzle using US data from the Survey of Professional Forecasters (SPF) and uncover evidence of over confidence in predictive distributions. Clements (2010) has compared the point predictions of professional forecasters with their subjective probability distributions, finding evidence that their means may be less frequently updated than their individual point forecasts. Also, Kenny, Kostka and Masera (2014) compare the density forecasts from the euro area SPF with a set of simple and naïve statistical models finding that the latter often outperform surveyed macroeconomic experts.

In this paper we exploit the microeconomic information contained in the European Central Bank (ECB) SPF as a means to contribute further to understanding the predictive performance of surveyed density forecasts.² In particular, we examine the possible role of density features such as their location, spread, skewness and tail risk in determining density forecast performance both over time and across individuals. Understanding this link is of relevance to both forecast users – such as the ECB - and forecast producers. In particular, the insights from our study are of interest to survey users who may rely on specific density features when evaluating different policy choices (e.g. relating to tail risks or possible skewness in the distributions). Moreover, density forecast producers, including those forecast producers responding to the SPF questionnaire, may improve their density forecast performance by gaining an understanding about how their density features have impacted on their historical density performance.

Our study thus contributes to a growing body of empirical work studying the properties of density forecasts collected in surveys. Diebold, Gunther and Tay (1998) and Diebold, Tay and Wallis (1999) evaluate the density forecasts from the US SPF while Casillas-Olvera

¹ For a survey of density forecasting see Tay and Wallis(2000) while Corradi and Swanson (2006) review methods to evaluate density forecasts.

² The micro dataset underlying our study can be downloaded from the ECB website at <http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html>. ECB (2014a) provides a recent overview of the survey drawing on the fifteen years of experience since its inception.

and Bessler (2006) and Boero, Smith and Wallis (2011) compare the density forecasts of Central Banks and professional forecasters. Our work is also closely linked to recent empirical work by Andrade, Ghysels and Idier (2012) who develop new SPF-based inflation risk indicators by extracting quantiles from individual surveyed distributions.³ Their results show, in particular, that these measures allow for a distinction between upside and downside risk which brings additional information to the consensus point forecast. A related strand in the literature has evaluated the probability forecasts for particular events (e.g. recessions, deflation) extracted from the surveyed SPF densities.⁴ A final strand of empirical work studies the surveyed density forecasts but focusses on the relationship between forecaster disagreement and uncertainty. In particular, in this strand of research, the surveyed histograms are used, either individually or in aggregate, to derive direct measures of uncertainty that can then be used to evaluate other ways of measuring or proxying forecast uncertainty. Such measures include the dispersion of individuals' point forecasts (a measure of disagreement). Relevant studies in this field are Zarnowitz and Lambros (1987), Giordani and Söderlind (2003), Boero, Smith and Wallis (2008), D'Amico and Orphanides (2008), Rich and Tracy (2010), Conflitti (2011) and Rich, Song and Tracey (2012).

Our analysis focuses on the one and two-year horizon density forecasts for euro area real output growth and consumer price inflation. We begin by constructing individual measures of density forecast performance from this dataset using the Ranked Probability Score (Epstein, 1969), recently advocated in Boero, Smith and Wallis (2011) as a preferred density scoring rule, especially for surveyed densities such as the one we examine. In particular, this performance measure is based on the entire predictive distribution – a feature which contrasts with the more commonly used (at least in economic applications) log predictive score. Next, we estimate directly at the individual level key moments of the SPF densities such as their means and higher moment features such as their variances, their skewness and tail probability mass. We then propose a set of cross-sectional and fixed effect panel regressions to examine the role of key

³ Garrat, Mitchell and Vahey (2013) combine the predictive densities from VAR models in order to derive probability forecasts that can quantitatively underpin central bank warnings about low inflation outcomes during the great recession. An interesting direction for future research may be the extent to which such model-based analysis can be combined with the corresponding indicators from surveyed densities such as those derived in the study by Andrade, Ghysels and Idier (2012).

⁴ For the US SPF, the relevant studies are Lahiri and Wang (2013), Lahiri and Wang (2007) and Clements (2006). For the euro area SPF, Kenny, Kostka and Masera (2013) have evaluated the derived event probabilities.

distributional features in explaining density forecast performance both across time and across individuals. Controlling for the role of differences in point forecast accuracy (density location) as well as for other common shocks impacting on aggregate density performance, these regressions help shed light on whether or not higher order density characteristics, such as variance, skewness or the fatness of a density forecast’s tails can contribute to improving forecast performance. Such a mode of analysis, responds to a clear need, as expressed in Leeper (2003) and Knüppel and Schulterfrankenfeld (2012), to generate empirical evidence concerning the quality of information that is contained in such density forecast features. Importantly, our analysis also sheds light on the use of judgement in forecasting and its impact on the quality of macroeconomic forecasts. In particular, in a recent questionnaire sent to the participants in the ECB SPF, a large majority of respondents (over 80%) indicated that their reported probability distributions are derived either purely on the basis of judgement or from models with judgemental adjustments.⁵

The layout of the remainder of the paper is as follows. In Section 2 we describe our approach to measuring density performance and estimating density moments. Section 3 outlines the econometric framework and estimation procedures we employ to study the link between density characteristics and density performance. Section 4 provides additional necessary background on the underlying data sources as well as some summary statistics on distributional properties of the SPF densities. Section 5 presents the main empirical findings on the link between density characteristics and predictive performance. Section 6 concludes.

2. MEASURING DENSITY PERFORMANCE AND MOMENTS

In this section, we present our main measure of individual density performance and discuss the methods we employ to estimate density moments such as the density mean, its variance, measures of skewness or the degree of probability mass in the tails of the distribution. In line with the discrete nature of the density forecasts from the SPF, we consider the probability forecasts, $f_{i,t+\tau}^k$, produced at time t of individual expert i ($i = 1, \dots, N$) defined over a finite set of outcome ranges (or “bins”) indexed from $k = 1, \dots, K^{\text{MAX}}$ for the forecast target variable of interest in period $t+\tau$, denoted $y_{t+\tau}$.⁶ $f_{i,t+\tau}^k$ thus

⁵ See Chart 10 in ECB (2014b).

⁶ In the ECB SPF dataset, as discussed in Bowles *et al.* (2010), the forecast horizon (τ) in each survey is set one and two-years ahead of the latest observed outcome and therefore differs across variables due to

represents the probability that expert i assigns to the event that the target variable will fall within the range covered by bin k . We index $k = 1, \dots, K^{\text{MAX}}$ where $K^{\text{MAX}} = \max \{K_t\}$ for $0 \leq t \leq T$, where T is the sample size and K_t denotes the number of bins used at time t which can change over time. In the ECB SPF, as in most other similar surveys, the intervals at the lower and upper end of the range of surveyed outcomes are open and thus capture the probabilities of all possible outcomes below a lower and an upper threshold. All other intervals are closed and capture the probability assigned to an interval of fixed width.⁷

2.1 Measuring density performance

A key element in our analysis is the definition of the appropriate loss function or rule, denoted by $L(\bullet)$, which measures the observed accuracy of the density forecasts. To do so, we follow Boero, Smith and Wallis (2011) who have recommended the Ranked Probability Score (RPS) due to Epstein (1969) as a preferred measure of density performance. In particular, in the present context, we favor the use of the RPS given its “sensitivity to distance” as discussed further below. Moreover, the predictive log score which has been used widely in economics is unsuited for individual surveyed densities given the high frequency of times when a zero probability is assigned to the actual outcome range for the predicted variable (see Kenny, Kostka and Masera, 2014).⁸

We use $x_{t+\tau}^k$ to denote the binary random variable taking a value of 1 if the period $t+\tau$ outcome occurs in “bin” k and zero otherwise. We can define the RPS for the density forecast in period $t+\tau$ in the case of the discrete probabilities associated with a survey such as the SPF as in equation (1) below.

$$L_{i,t+\tau} = \sum_{k=1}^{K^{\text{MAX}}} \left(F_{i,t+\tau}^k - X_{t+\tau}^k \right)^2 \quad (1)$$

differing publication lags for the release of official HICP and GDP statistics. For example, the one-year-ahead GDP forecast refers to the annual growth rate two quarters after the survey quarter, whilst the equivalent HICP forecast refers to the annual inflation rate approximately 11 months after the survey month. For notational convenience, we nonetheless refer to these variable 1 and 2 year ahead rolling horizon forecasts as H=1 and H=2 respectively.

⁷ In our sample the maximum number of bins used has differed for GDP and inflation variables, with the former comprising 24 bins ranging from -6.0 to +4.9 and the latter comprising 14 bins ranging from -2.0 to +3.9. The width of the bins in the ECB SPF has been held constant at 0.4 for both GDP growth and inflation.

⁸ Aside from this practical problem confronting its use with survey data, the log score is a valid and attractive measure of density performance. In particular it is also a strictly proper scoring rule.

In (1), $F_{i,t+\tau}^k$ and $X_{i,t+\tau}^k$ denote the cumulative distribution functions of the surveyed densities and the binary outcome variable respectively. As it is based on the cumulative distributions, the RPS will tend to penalize less severely density forecasts which assign relatively larger probabilities to outcomes that are close to the outcome range that actually occurs. In this sense, the RPS is “sensitive to distance” in contrast to, for example, a *local* scoring rule such as the log score that only ranks densities on the basis of the probability they assign to the outcome that occurs. The RPS is bounded over the interval $[0, K^{MAX} - 1]$, reaching a minimum value of 0 only in the extreme circumstances when all the probability is placed in a single bin and the actual outcome falls precisely into that bin. Correspondingly, its maximum value is obtained when a unit probability is assigned to an interval at one end of the range of possible outcomes but the outcome occurs at the opposite end of the range of possible outcomes. Reflecting its dependence on the cumulative distributions and given the unpredictability of economic events, i.e. forecasters do not have perfect foresight, the expected value of the RPS will be positive, i.e. $E[L_{i,t+\tau}] \geq 0$. Also, as a strictly proper scoring rule (see Gneiting and Raftery, 2007), the RPS provides a penalty for density misspecification in a way that elicits forecasters’ true beliefs over time.⁹ Given its sensitivity to distance, the RPS gives some relative reward to a density forecast that has a “near miss” relative to one that assigns higher probabilities in regions very far away from the actual outcome. This sensitivity of the RPS to distance is particularly appealing for analyzing the impact of density features on overall forecast performance. For example, for a given probability mass assigned to the outcome, the RPS will reward higher moment features such as skewness or thickness of the tails if such features give rise to greater probability mass in regions that are close to where the outcome occurs.

One issue that arises when seeking to compute the RPS using SPF data is how to treat the open intervals at the extremities of the surveyed histograms. Most studies (e.g. Zarnowitz and Lambros, 1987, D’Amico and Orphanides, 2008) have suggested equating the open intervals with an equivalent closed interval of equal width. However, if the probability in the open intervals is relatively large, the assumption that it is all concentrated in a single closed interval may be less justified. Therefore we adopt a modified approach of assigning all the probability to a single closed interval of equal width when the reported probabilities are very small and below 1.0%. For cases where the probability is above

⁹ As discussed in Diebold, Gunther and Tay (1998) “there is no way to rank two incorrect density forecasts such that all users will agree with the ranking” (p. 866). While the RPS provides an attractive and meaningful ranking of the densities, particularly for the type of surveyed data that we employ, our study is still to a largely conditional on this choice.

1.0%, we assign 2/3 of it to an equivalent closed interval and the remaining 1/3 to the next lower or higher interval depending on which side of the distribution one is working with. For example, if 0.5% probability was assigned to an open interval, all of this is taken as being assigned to an equivalent closed interval. However, for a larger probability of say 2.0% assigned to an open interval, e.g. ≤ 0.0 , 1.33% would be assigned to a closed interval $[-0.4, 0.0]$ and 0.66% would be assigned to $[-0.8, -0.4]$. This approach has the desirable feature that intervals that are further away from the center of the distribution have a lower probability mass.¹⁰ In practice, the vast majority of forecasters tend not to assign *any* probability to the open intervals and, hence, the analysis is largely unaffected by which of these approaches are used.

A second practical issue that arises when computing the RPS for a given density forecast is the choice of vintage for the outcome variable that is used to compute the interval into which the outcome falls and hence the values taken by $X_{t+\tau}^*$ in equation (1). For example, one could define the true outcome using a very recent vintage or alternatively one could opt for a more real time vintage that reflects the initial statistical release. This choice could impact the performance ranking of individual densities especially when there are substantial revisions in the data. For example, Genre, Meyler, Kenny and Timmermann (2013) find some notable effects of data revisions in the evaluation of SPF forecasts for the unemployment rate. However, for inflation and GDP growth, such revisions have been more marginal and forecast evaluation appears less sensitive to the choice of data vintage. In what follows, therefore, we report results only using the recent (i.e. 2013 Q1) data vintage for both variables.¹¹

Figure 1 plots the RPS measures for the GDP growth and HICP inflation densities from the euro area SPF over the period 1999Q1-2013Q1. The plots show for both 1 year and 2 year density forecasts ($H=1$ and $H=2$ respectively) the RPS measure for the median forecaster. In addition, the performance at the extremities is represented by the 10th and 90th percentiles of the RPS distribution. By this measure we can observe that, for both variables, density forecast performance has varied considerably over the sample period. For both GDP growth and inflation, the median performance deteriorated substantially as indicated by exceptionally elevated scores during the 2008-2009 periods which were

¹⁰ We have also checked the sensitivity of our analysis to other possible assumptions of equating the open intervals with one closed interval or spreading the probability uniformly over two closed intervals. However, reflecting the relatively low utilization of open intervals in the panel dataset, this has no noticeable impact on the main empirical results we report in Section 5.

¹¹ However, we have also recomputed all our results using the 1st available vintage for both GDP and inflation. As expected, we find no noticeable impact on our main findings.

significantly influenced by the financial crisis.¹² For both variables, but particularly for inflation, there is some evidence that the predictive performance of the SPF densities was worse during the initial years of the survey (2000-2003) than it was in the period prior to the crisis (2004-2007). This gradual improvement is suggestive of possible learning behavior on the part of forecasters. Given that the euro area was a new economic entity which had not been the focus of macroeconomic attention prior to the launch of the single currency in 1999, some learning on the part of forecasters could be expected. As represented by the 10th and 90th percentiles from the cross section of RPS scores, some notable variation in predictive performance across individual forecasters has also been observed. For GDP growth, such heterogeneity is most evident during the period least influenced by the crisis, while during the crisis years the 10th and 90th percentiles are very close to each other implying all forecasters tended to do equally poorly. For HICP inflation, some noticeable heterogeneity appears to persist even during the years influenced by the crisis.

2.2 Density moments

As noted above, the surveyed density forecasts from the SPF are represented by discrete distributions whereby survey respondents assign a particular probability mass to an outcome range for the forecast target variable. One approach to estimating density moments is to fit continuous densities to the individual surveyed histograms as in, for example, Engleberg, Manski and Williams (2009). Moreover, it is not self-evident that the survey replies should be interpreted as being drawn from a deeper subjective continuous distribution that is present in the minds of respondents. Moreover, such an approach entails imposing some distributional assumptions on the surveyed densities which may pose practical challenges given that individual histograms are often restricted to very few bins. As an alternative, and in line with the earlier work of Lahiri, Teigland and Zaporowski (1988), we calculate density moments using the discrete approximations which correspond closely to the discrete histograms included in the survey

¹² A close inspection of the SPF data from the 2009Q1 survey shows a piling up of probability mass in the interval $<-1.0\%$ for the GDP forecasts while the reported point forecasts were located well below this interval reflecting the strength of the deterioration in the euro area outlook. Such a piling up of probability mass suggests that respondents may have been unable to report their true subjective densities because they were not supplied with a sufficient range of outcomes in the survey questionnaire. We therefore conducted all our analysis also excluding the forecasts from the 2009Q1 survey. Overall we find very similar results, suggesting that the measurement error associated with this survey round has not unduly influenced our findings.

questionnaires.¹³ To do so, we make the assumption that all the probability mass in a given bin k is assigned to the midpoint of that bin (μ_k). The estimated moments are then computed as probability weighted averages given in (2) to (5) below.¹⁴

Mean

$$\mu_{i,t+\tau} = \sum_{k=1}^{K^{MAX}} f_{i,t+\tau}^k \mu_k \quad (2)$$

Variance

$$\sigma_{i,t+\tau}^2 = \sum_{k=1}^{K^{MAX}} f_{i,t+\tau}^k [\mu_k - \mu_{i,t+\tau}]^2 \quad (3)$$

Absolute skewness¹⁵

$$\gamma_{i,t+\tau} = \left| \frac{\sum_{k=1}^{K^{MAX}} f_{i,t+\tau}^k [\mu_k - \mu_{i,t+\tau}]^3}{\sigma_{i,t+\tau}^3} \right| \quad (4)$$

Excess kurtosis

$$\zeta_{i,t+\tau} = \frac{\sum_{k=1}^{K^{MAX}} f_{i,t+\tau}^k [\mu_k - \mu_{i,t+\tau}]^4}{\sigma_{i,t+\tau}^4} - 3 \quad (5)$$

To illustrate our estimates, Figures 2 and 3 plot the above four moments from the cross section of individual densities for euro area GDP growth and HICP inflation,

¹³ See also Giordani and Söderlind (2003) and Conflitti (2011) for other application of discrete approximations to surveyed distributions.

¹⁴ In Section 3, we propose a panel regression linking density performance with these estimated moments. Given the need to estimate density moments from the underlying survey data, it is important to consider the potential complications for our subsequent empirical analysis that can arise as a result of measurement error (see, for example, Curtin, 2010)). In particular, following Grilliche and Hausmann (1986) we control for stable sources of measurement error and check the robustness of our results by comparing both levels and first differences of the moments in the regressions.

¹⁵ In our empirical analysis, we primarily focus on the absolute skewness rather than the actual skewness in order to focus attention on the hypothesis that more skewed distributions are associated with better density forecast performance (i.e. lower scores). However, in Figures 2 and 3 discussed below, we also report actual skewness and discuss the extent of positive and negative asymmetries in the SPF distributions.

respectively, at $H=1$.¹⁶ As in Figure 1, we illustrate the cross sectional distribution in SPF density forecasts by plotting the median value for each moment together with the 10th and 90th percentile values. Both Figures point to some notable cross sectional variation in density properties. For example, during the depth of the recent recession in 2009, mean forecasts for the annual change in GDP growth varied between -3.0% (10th percentile) and close to zero (90th percentile). In terms of the variances of the individual SPF densities, there is also considerable heterogeneity. As reflected in a persistently low variance at the 10th percentile, some forecasters have tended to exhibit excessively high confidence in their mean predictions. For both inflation and GDP growth, variances also tended to rise following the recent recession and, although they have subsequently declined, for many individuals they have yet to return to pre-crisis levels.¹⁷

In terms of the two other higher moment features, we also observe considerable variation over time and across forecasters. For both GDP growth and inflation, while the median forecaster has often tended to construct a relatively symmetric density forecast, throughout the whole sample period there is evidence of other forecasters with both positive and negative skewness in their densities. Lastly, the estimates of negative kurtosis for the median SPF density forecast for both inflation and GDP growth are indicative of relatively low tail risk (e.g. relative to a Gaussian distribution with the same variance). Another interesting feature that emerges from Figures 2 and 3 is that, in contrast to the means of the densities where the 10th and 90th percentiles tend to strongly move together over time, for higher moment features such strong co-movement is less prevalent. In other words, there is evidence of more heterogeneity in the movement of these higher moment features over time. In our subsequent analysis, we investigate the impact of this cross sectional and time variation in higher moment features on overall density forecast performance.

3. LINKING PERFORMANCE AND DENSITY CHARACTERISTICS

In this section we describe the econometric framework for the analysis of the link between density features and density forecast performance. To consider the relationship

¹⁶ The equivalent figures for $H=2$ provide a very similar impression.

¹⁷ The estimated variance provides a measure of *ex ante* uncertainty as perceived by SPF forecasters. In a recent study, Clements (2012) has examined the relationship between *ex ante* uncertainty and *ex post* or realized uncertainty.

between individual density performance and key distributional properties, we propose a panel regression model of the form given by (6).

$$L_{i,t+\tau} = \alpha_i + Z_{i,t+\tau}\beta + \varepsilon_{i,t+\tau} \quad (6)$$

In (6) $L_{i,t+\tau}$ refers to the performance measure (RPS) of forecaster i given the outcome for the target variable in period $t+\tau$ and $Z_{i,t+\tau}$ is a vector of regressors capturing density features such as location, variance, skewness and kurtosis that may be embodied in the subjective probability forecasts. The vector of estimated parameters (β) captures the sensitivity of individual density performance to density characteristics. As (6) captures the impact of variation in distributional features over time, it is important to control for other factors which may vary over time but which are fixed across forecasters. Such factors include, for example, common sources of business cycle fluctuations or aggregate macroeconomic shocks. We therefore estimate equation (6) using time fixed effects (α_i) to capture the sources of density forecast performance that are common to all forecasters.¹⁸ The use of fixed effects to capture aggregate shocks follows previous panel applications to such surveyed data for the case of point forecasts (see, for example, Davies and Lahiri (1995), Davies and Lahiri (1999) and Boero, Smith and Wallis (2008)). As described below, for the panel model, we use a Feasible Generalized Least Squares (FGLS) estimation procedure which takes account of any off-diagonal components of the residual variance covariance matrix in (6).

A key advantage of the specification in (6) is that it provides a relatively parsimonious representation of the link between density performance and density characteristics but it identifies jointly the role of key distributional features, rather than focusing partially on any one feature. As highlighted in Section 2, we do not expect the RPS to be zero, i.e. $E[L_{i,t+\tau}] \geq 0$. Rather it can provide a metric of performance over time and across forecasters that can be empirically linked to the measured characteristics of the individual densities. As a strictly proper scoring rule the RPS provides an estimate of how far away a given density forecast is from the true data generating process. Clearly, one would expect that a large share of a given density's performance will be determined by its

¹⁸ We have also considered the GMM estimator of which FGLS is a special case and found very similar results. Also, as an alternative way to control for the impact of aggregate shocks, we considered the use of a group mean estimator using the Common Correlated Effects for Pooled regressions (CCEP estimator) due to Pesaran (2006). We do this by augmenting the estimation regression with cross-section averages of independent and dependent variables and found very similar results.

location, i.e. the mean forecast of the expert relative to the outcome.¹⁹ Such a perspective justifies the inclusion of the mean absolute error on the right hand side of the panel regression as it provides an estimate of the difference between the true mean and the mean of the expert density. Clearly, one would anticipate a positive relationship between mean forecast accuracy and density performance. Controlling for this effect, however, in investigating the relative degree of skill embodied in the overall predictive distribution, it is of interest to consider the impact on the overall density score of distributional features such as the variance, the skewness or the degree of probability mass allocated to the tails. For these higher order features we do not observe their true values or outcomes and, hence, it is not possible to propose a specification in terms of deviation from true values where again we would expect a positive relationship, e.g. with excessive variance, skewness or kurtosis being penalised and thus resulting in worse performance.

In the remainder of this section, we discuss in more detail the expected sign of the coefficients on these higher moment characteristics and we also propose tests of the robustness of the empirical findings along various dimensions including estimation in first differences and also by considering a nonlinear representation of the panel model where various asymmetries are taken into account.²⁰

3.1 Cross-sectional regression

A first perspective on the relationship between distributional characteristics and density forecast performance is to consider only the evidence in the cross section, averaging over

¹⁹ Throughout the paper, we measure point forecast accuracy using the estimated mean of the density forecast as estimated by equation (2). However, in the SPF survey, in addition to the reported density forecasts, respondents also report “point forecasts” for equivalent variables and horizons and these can differ from the estimated means. At the suggestion of an anonymous referee, we have therefore checked the sensitivity of our results to this choice and recomputed all our empirical analysis but using these point forecasts. Overall, we do not find that this choice has impacted the main conclusions we derive. This is consistent with a methodological survey of the ECB SPF conducted in 2009 in which a large majority of participants reported that their point forecasts correspond to the means of their reported densities. The results of this methodological survey are available at:

<http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html>

²⁰ Our approach has some parallels with the work of Mitchell and Wallis (2011) who focus on density evaluation with probability integral transforms (pits). They suggest regressing the pits on explanatory variables in order to gain insights into the factors that might explain good and bad performance. Indeed, more generally, inspection of the pits can provide insights into the sources of density forecast performance such as a too high or low mean/variance or an excessively small/high tail probability mass relative to the true data generating process.

time the dependent and independent variables in (6). This suggests a cross sectional regression of the form given in (7) below.

$$L_i = \alpha + \beta_1 |y - \mu_i| + \beta_2 \sigma_i^2 + \beta_3 \gamma_i + \beta_4 \zeta_i + \varepsilon_i \quad (7)$$

In equation (7), L_i refers to the sample average score of each individual density forecast, $|y - \mu_i|$ to the corresponding individual sample mean absolute forecast error, σ_i^2 to the sample average variance of forecaster i 's density forecasts, γ_i to the sample average *absolute* skewness, and ζ_i to the sample average level of kurtosis. Equation (7) thus offers some insight into understanding the relevance of distributional features in systematically explaining density forecast performance across forecasters. *A priori* one would anticipate $\beta_1 > 0.0$, i.e. forecasters with relatively large mean absolute forecast errors have relatively poor density forecast performance (higher scores). The remaining parameters will tend to highlight any systematic differences across forecasters in the other distributional features (i.e. not related to point forecast accuracy) but which nonetheless impact on density performance. For example, to the extent that some forecasters tend to be overconfident or neglect certain important risks which materialized over the sample, they will tend to have low variances but relatively high scores. Conversely, those who least neglect these important risks (having higher variances) may be able to improve their overall density performance, i.e. lower their density scores compared with other low variance forecasters. Under these conditions, one would anticipate a significant negative relationship between average individual variance and average individual score. Similarly, the cross-sectional regression allows us determine whether individuals with more skewed distributions or distributions with greater tail mass perform better than those where such features are less prevalent. This might arise if, for example, some forecasters have neglected important skewness or tail risks (such as those which materialized in 2008-2009) while others do not or at least neglect them to a lesser extent.²¹

²¹ Neglecting risks is not problematic *per se*. For example, the literature on rational inattention (see Sims, 2003) has demonstrated that it may not be optimal to plan for rare (low probability) events such as earthquakes and financial disaster. Within our framework, it is the neglect of risks that have some significant chance of occurring that can be associated with a worse density performance. In a forecasting context, neglecting the possibility of certain outcomes may also be fully rational if those outcomes are not feasible or possible given the underlying economic reality. However, assigning a zero probability to outcomes that have some chance of occurring will tend to be penalized by strictly proper scoring rules such as the log score or the RPS. Indeed for the log score, the penalty for giving zero probability to an event that subsequently materializes is infinite and as a result a forecaster seeking to maximise performance should only assign a zero probability to an event or outcome that is truly impossible (e.g. unemployment rates that are less than 0%). In the case of the discrete distributions of the SPF, this would

3.2. Panel regression

Another perspective on the link between distributional characteristics and density forecast performance can be obtained by estimating a full panel version of (7). The panel analogue of equation (7) is given in (8).

$$L_{i,t+\tau} = \alpha + \beta_0 D_t + \beta_1 |y_{i,t+\tau} - \mu_{i,t+\tau}| + \beta_2 \sigma_{i,t+\tau}^2 + \beta_3 \gamma_{i,t+\tau} + \beta_4 \zeta_{i,t+\tau} + \varepsilon_{i,t+\tau} \quad (8)$$

One important feature of (8) is that the sensitivity of each individual's density performance to individual characteristics is assumed to be constant across individual forecasters. This pooling of the regression coefficients strikes us as reasonable. It is hard to think of reasons why a given individual's performance would be more sensitive to fluctuations in a given characteristic compared with any other individual. At the same time, the commonality of the coefficients is directly testable given that we have a sufficiently large number of time series observations. These tests, which we refer to briefly in Section 5, overwhelmingly support the assumption of fixed coefficients across forecasters in the panel. Importantly this does not imply that all individuals are the same but rather that the heterogeneity in forecast performance can be fully explained by a common response to each individual's density characteristics.

As mentioned previously, given that (8) does not average performance across time, it is important to control for the role of common shocks across individuals. We therefore estimate equation (8) using a T-dimensional vector of dummy variables (D_t) for time fixed effects to capture common sources of business cycle fluctuation or aggregate macroeconomic shocks (i.e. $\alpha_t = \alpha + \beta_0 D_t$). Another likely factor that may impact the efficiency of the estimates in our panel regression is the possibility of serial correlation in the errors such that $E[\varepsilon_{i,t}, \varepsilon_{i,t+j}] \neq 0$ for $j = 1, \dots, \tau$ and $i = 1, \dots, N$. Hence, we also control for possible correlation in the errors of the panel regression that may result from the multi-period nature of the forecast horizon and the quarterly frequency of the survey and estimate (8) using a Feasible Generalized Least Squares (FGLS) procedure (See Wooldridge, 2002). Under this procedure, and using the notation of equation (6), we first estimate the equation residuals by estimating (8) by OLS. From the residuals of the OLS regression, we then derive estimates of the error correlation and construct a new estimate of the error variance covariance matrix ($\hat{\Omega}$) for our panel regression allowing for auto

tend to imply always having some very small positive probability in any outcome ranges that are technically feasible. In the case of the RPS, as highlighted in Section 2, the penalty for neglecting risks that materialize is bounded given that the RPS itself is bounded.

correlation of up to order τ .²² The estimated parameters of the panel regression and their associated standard errors are then given by $\hat{\beta}_{FGLS} = (Z' \hat{\Omega}^{-1} Z)^{-1} (Z' \hat{\Omega}^{-1} L)$ and $Var[\hat{\beta}_{FGLS}] = (Z' \hat{\Omega}^{-1} Z)^{-1}$.²³

Lastly, given the possible impact of measurement error on our results, associated with the estimation of density moments, we also report the estimation of equation (8) in first differences. This follows the suggestion in Grilliche and Hausmann (1986) to check for any influence of stable sources of measurement error by comparing the panel regressions in levels and first differences.²⁴ Using $\Delta L_{i,t+\tau}$ to represent the change in the density score between period $t+\tau$ and $t+\tau-1$, the first difference specification is denoted by (9) below.

$$\Delta L_{i,t+\tau} = \alpha + \beta_0 D_t + \beta_1 \Delta |y_{i,t+\tau} - \mu_{i,t+\tau}| + \beta_2 \Delta \sigma_{i,t+\tau}^2 + \beta_3 \Delta \gamma_{i,t+\tau} + \beta_4 \Delta \zeta_{i,t+\tau} + \varepsilon_{i,t+\tau} \quad (9)$$

3.3 Panel regression with asymmetries

The empirical relations described by equations (8) and (9) assume a symmetric and linear relation between density performance and density characteristics. However, there is a case to consider a possible nonlinear relation between density features and density characteristics. Likely sources of nonlinearity are potential asymmetries such that positive changes in density characteristics may have a different impact on performance compared with negative differences.

An asymmetric panel regression in first differences, also controlling for factors which are common across forecasters using time dummies, may help better identify individual forecaster skill. For example, to the extent that the variance of the true density has varied

²² In practice autocorrelations greater than 4 tend to be very small and can be discarded in the FGLS procedure. For the one-year-ahead horizon this is entirely consistent with what would be expected for a four quarter ahead forecast sampled at a quarterly frequency but it is more surprising for the two year ahead forecast.

²³ In the estimation of the adjusted error variance matrix, we also considered possible correlation in the errors across individual forecasters that might arise due to common aggregate shocks in line with the panel analysis of point forecasts in Keane and Runkle (1990). However, we did not find these correlations to be important. A likely explanation for this finding is that the impact of common shocks is adequately captured through the inclusion of the fixed effects time dummies.

²⁴ An alternative would be the consideration of estimation using instrumental variables but we did not pursue because, aside from a lack of any obvious choice of instruments for density moments that we are using, the measured correlation between the regressors and the estimated residuals was always very close to zero.

over time, e.g. as a result of increasing or decreasing macroeconomic uncertainty, skilled macroeconomic experts would be able to improve their density forecast performance by varying their predictive variances in line with this. Such skill would imply that both positive and negative changes in these distributional features could result in an improvement in density forecast performance, as reflected in lower density forecast scores. In the light of these arguments and, as an additional robustness check, we estimate a version of the panel regression in first differences and consider separately the effect of increases and decreases in higher moments.²⁵ The resulting specification is given in equation (10) below.

$$\begin{aligned} \Delta L_{i,t+\tau} = & \alpha + \beta_0 D_t + \beta_1 \Delta |y_{t+\tau} - \mu_{i,t+\tau}| + \beta_2 (\Delta \sigma_{i,t+\tau}^2)^+ + \beta_3 (\Delta \sigma_{i,t+\tau}^2)^- + \beta_4 \Delta \gamma_{i,t+\tau}^+ + \beta_5 \Delta \gamma_{i,t+\tau}^- \\ & + \beta_6 \Delta \zeta_{i,t+\tau}^+ + \beta_7 \Delta \zeta_{i,t+\tau}^- + \varepsilon_{i,t+\tau} \end{aligned} \quad (10)$$

In (10) $\Delta L_{i,t+\tau}$ again represents the change in the density score between period $t+\tau$ and $t+\tau-1$. However, in contrast to the specification in (9), for each of the regressors corresponding to this, we construct a dichotomous variable which takes on the value of the distributional characteristic depending on whether the change in that characteristic is positive or negative and takes a value of zero otherwise, i.e.

$$(\Delta \sigma_{i,t+\tau}^2)^+ = \begin{cases} \Delta \sigma_{i,t+\tau}^2 & \text{if } \Delta \sigma_{i,t+\tau}^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$(\Delta \sigma_{i,t+\tau}^2)^- = \begin{cases} \Delta \sigma_{i,t+\tau}^2 & \text{if } \Delta \sigma_{i,t+\tau}^2 < 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta \gamma_{i,t+\tau}^+ = \begin{cases} \Delta \gamma_{i,t+\tau} & \text{if } \Delta \gamma_{i,t+\tau} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta \gamma_{i,t+\tau}^- = \begin{cases} \Delta \gamma_{i,t+\tau} & \text{if } \Delta \gamma_{i,t+\tau} < 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta \zeta_{i,t+\tau}^+ = \begin{cases} \Delta \zeta_{i,t+\tau} & \text{if } \Delta \zeta_{i,t+\tau} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta \zeta_{i,t+\tau}^- = \begin{cases} \Delta \zeta_{i,t+\tau} & \text{if } \Delta \zeta_{i,t+\tau} < 0 \\ 0 & \text{otherwise} \end{cases}$$

²⁵ We estimate the regression with asymmetries only in first differences given that the variance is always positive in levels whilst in a first difference specification we are able to distinguish between positive and negative changes in the variance.

Estimation of (10) thus reveals any asymmetry in the impact of density features, distinguishing between periods when the overall dispersion, the absolute skewness or the excess kurtosis is either increasing as opposed to decreasing. For example, any asymmetry in the impact of increases in variance compared with decreases in variance can be revealed by any estimated difference in the estimated values for β_2 compared with β_3 . Similarly, any asymmetries in relation to skewness and excess kurtosis are revealed by a comparison of the estimated values of β_4 with β_5 and of β_6 with β_7 , respectively.

4. SPF DENSITIES: DATA AND SUMMARY STATISTICS

The micro dataset underlying our study can be downloaded from the ECB website at <http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html>. Our analysis is based on the one and two-year horizon density forecasts for euro area real output growth and consumer price inflation, with the outcomes for these variables being measured, respectively, by euro area real Gross Domestic Product (GDP) and the Harmonised Index of Consumer Prices (HICP) as published by Eurostat, the statistical agency of the European Union. The SPF dataset comprises a highly unbalanced panel reflecting the irregular responses of respondents to the survey. In order to limit the impact that such sampling issues may have on our results, following a common strategy in empirical studies with this data (see Kenny, Kostka and Masera, 2014), we have filtered the data to include only regular respondents. The latter are defined to be those respondents with no more than four missing replies in a row. Such filtering ensures that our findings are based on a consistent set of time periods that are shared across the forecasters that we analyze. However, it comes at a cost as there is a considerable reduction in the number of individuals included in our analysis from between 35 and 50 replies on average in the unfiltered panel to 25 and 24 individuals for the one- and two-year horizons, respectively. In addition, we exploit the observed persistence in individual density features in order to construct a fully balanced panel of density forecasts from this filtered dataset.²⁶

In the remainder of this section we summarize the main features of the SPF densities and explore in particular the degree of heterogeneity in the characteristics of the densities

²⁶ The approach to balancing the panel builds on Genre, Kenny, Meyler and Timmerman (2013) which uses a pooled regression to measure the degree of persistence in the deviation of an individual point forecasts from the average forecasts. This provides an update for the location of the density. The updated density is then centred on the bin containing this updated forecast and the associated probabilities are derived using the most recently observed probabilities that were submitted. Kenny, Kostka and Masera, 2013, p. 17-19, provide a more complete description of this procedure.

across individuals. Importantly, SPF replies are collected on an anonymous basis – a feature which may help limit the role of strategic incentives in driving the observed differences across forecasters. Tables 1 and 2 report for the one and two year ahead ($H=1$ and $H=2$ respectively) GDP growth and inflation density forecasts, the first four moments of the individual densities calculated as described in Section 2. In order to focus only on the cross sectional pattern, each distributional characteristic is reported as a sample average taken over the period 1999Q1-2013Q1. The estimated moments are then reported for the cross section of individual replies as summarised by the Median (50th percentile), the Max/Min, the 10th/90th and 25th/75th percentiles of the distribution. To complement the information on these distributional features, the tables also report the Mean Error (ME) and the Mean Absolute Error (MAE) for the cross section of individual forecasters, calculated using the estimated mean forecasts taken from equation (2).

Surveying the summary statistics in Tables 1 and 2 highlights positive mean errors for GDP and negative mean errors for inflation, indicating a tendency for the individual density forecasts to over predict GDP growth and under predict inflation during the sample period. The tables also display considerable heterogeneity across forecasters in terms of mean forecast accuracy. For example, at $H=1$, the individual with the most accurate mean forecast for GDP growth has a MAE that is close to 30% better than the worst performing forecasters. Similarly, although GDP growth forecasts tend to over predict growth on average, some forecasters (close to 25% of them), have average errors that are below 0.2 percentage points (p.pt.). A similar degree of forecaster heterogeneity is observed for the individual inflation forecasts. Turning to the other sample statistics highlighted in Tables 1 and 2, a very clear feature of the surveyed replies is the low variance of the expert density forecasts, in particular relative to the MAEs. For example, although the median forecaster has a MAE of 1.20 p.pt for GDP growth at $H=1$, the median variance is only 0.20 p.pt. Even the individual with the highest spread in his predictive distribution (0.79 p.pt.) still has a variance that is below the MAE of the best performing mean forecast. Such a mismatch between the variances and the out-of-sample point forecast accuracy in the expert forecasts is suggestive of considerable overconfidence on the part of macroeconomic experts. It is observed at both forecasting horizons and for both GDP growth and inflation, although for the latter variable the mismatch is more modest because realized uncertainty as measured by the MAE has been lower for inflation. This evidence is in line with previous studies such as Giordani and Söderlind (2006) and Clements (2012) for the US and Kenny, Kostka, Masera (2014) for the euro area. It is also in line with experimental evidence of expert overconfidence from behavioral economics and psychology (Kahneman and Tversky, 2000). Finally, it is

notable that the variances of the surveyed densities are only marginally increasing in the forecast horizon, suggesting that macroeconomic experts are almost as confident about their two-year-ahead predictions as they are about their one-year predictions.

The cross section of individual skewness and kurtosis measures displayed in Tables 1 and 2 also highlights considerable variation across forecasters (as was also documented in the discussions of Figures 2 and 3). The individual surveyed densities generally exhibit positive absolute skewness, confirming that experts use their density forecasts to express a directional view concerning the overall balance of risks around their mean predictions. The cross sectional pattern of the absolute skewness tends to be quite consistent across forecast horizons and variables, with the most skewed distributions being approximately three times more skewed than the density with median skewness. As documented previously in relation to Figures 2 and 3, there is a strong tendency for individual forecasters to display negative excess kurtosis, i.e. a lower probability mass in the tails of their distributions relative to a Gaussian distribution with the same variance. For both GDP and inflation, a few individual forecasters attribute on average higher tail probabilities than an equivalent Gaussian distribution. However, positive excess kurtosis is more the exception. This finding of relatively low tail risks in SPF distributions at the individual level is once again in line with the relatively high level of confidence embedded in these expert density forecasts.

5. EMPIRICAL RESULTS

In this section we discuss the empirical evidence on the link between individual density characteristics and overall density forecast performance. We focus first on the cross sectional evidence from the estimation of (7), averaging density performance and density features over time and then turn to the panel analysis described by equations (8), (9) and (10).

5.1 Cross sectional evidence

The results from the estimation of equation (7) are reported in Table 3a, with the cross sectional analysis comprising 24 and 25 individuals for, respectively, one- and two-year horizons. Overall, the cross sectional regression tends to explain a large fraction of the variation in density performance across individuals with, depending on the horizon and the variable, adjusted R^2 that range between 90% and 96%. The coefficient estimates also yield a number of clear insights. As expected, experts with less accurate mean forecasts in general also have less accurate density forecasts as indicated by a significant and positive

β_1 for each variable and at each horizon. Another systematic feature contained in the cross-section regression is a sizeable and statistically significant negative coefficient (β_2) on the variance. Across the SPF panel, a higher variance is associated with lower scores and, hence, better performance. This suggests that some density forecasters in the panel with higher average variances systematically perform better than others implying an important source of heterogeneity in the density replies. To the extent that the scoring rule used penalizes symmetrically positive and negative deviations from the true but unknown variance, and if some forecasters underestimated variance while others overestimated it to a similar degree (i.e. there was no tendency to underestimate the variance on average), one would not anticipate any systematic relationship between density score and density variance in the cross-section. Hence, the finding of a systematic negative relation between density performance and the variance, which is observed for both inflation and growth and at both horizons, is very much in line with the evidence of overconfidence and neglected risks by many experts in the panel as highlighted previously in the discussion of Tables 1 and 2.

The above findings suggest that forecasters with relatively high (low) variance tend to have systematically better (worse) density forecast performance. Thus many forecasters in the panel are operating in a region where density performance can be improved by increasing the spread of their reported distributions. More generally, the significant coefficients in the cross sectional regression points to some notable heterogeneity in both the average *ex post* performance and the *ex ante* characteristics of the densities of the professional forecasters of the ECB SPF. This is in line with recent evidence in Boero, Smith and Wallis (2012). Using the Bank of England’s survey of external forecasters, they find substantial heterogeneity in forecasters’ uncertainty about future outcomes, as expressed in their subjective probabilities, and strong persistence in the relative level of individual forecasters’ uncertainty.²⁷

Looking in more detail at the results in Table 3a, the cross section regression suggests the negative impact of variance is somewhat stronger for the short horizon density forecasts and is particularly evident for GDP growth. For inflation, the estimated β_2 parameters are smaller in absolute terms which suggests such overconfidence is perhaps quantitatively less important but nonetheless significant (in line with our previous observations in

²⁷ Of course, heterogeneity in *ex post* performance does not imply it is easy to identify good forecasters *ex ante*. In line with this, for the case of point forecasts using the US SPF, D’Agostino, McQuinn and Whelan (2012) find limited evidence for the idea that the best forecasters are “innately” better than others.

relation to Table 2).²⁸ Turning to the remaining parameter estimates, we find that forecasters with more skewed distributions in absolute terms tend – if anything - to do worse on average. This is indicated by an estimated positive value for β_3 which is significant for the GDP densities at H=1. Such a finding would cast some doubt on the “information value” of the skewness assessments embodied in expert predictions, a result which compares closely with results in Knüppel and Schulterfrankenfeld (2012) who report no conclusive evidence for a systematic connection between risk assessments of central banks and their forecast errors. Indeed, for growth forecasts at the shorter horizon our results document a statistically significant deterioration of performance for experts with more skewed distributions compared with that of experts with less skewed distributions. Lastly, although increased probability mass in the tails is sometimes associated with an improvement in individual performance ($\beta_4 < 0.0$), this effect is quantitatively small and never statistically significant. The forecasters with higher probability mass in the tails of their distributions do not achieve a systematic improvement in predictive performance compared with other forecasters.

The above finding on the negative relationship between variance and density performance is one of the paper’s key results. Therefore it is of interest to also check the robustness with respect to alternative measures of the dispersion in the expert densities. Rich, Song and Tracey (2012) have suggested the Inter Quartile Range (IQR) as providing a robust measure of individual uncertainty as it will tend to be less influenced by outliers and also to other sources of measurement error. Hence, we have therefore recomputed all our regressions using this measure and the results are reported in Table 3b). Once again this new specification points to a strong negative relationship in the cross section, between the spread of the distribution as measured by the IQR and its overall predictive accuracy. This holds across both variables and both horizons, while it is particularly strong for the GDP densities, a result which mirrors also the findings when the estimated variance is used.²⁹ The significant worsening impact of increased skewness for GDP (H=1) densities is, however, not retained in the regressions using the IQR.

²⁸ Given the dependence of the RPS performance metric on the maximum number of bins used, and given that this differs for the GDP growth and inflation densities, some caution is warranted in drawing strong inference based on the size of the estimated coefficients.

²⁹ The relative insensitivity of our findings with respect to the choice of how the density spread is measured also tends to rule out measurement error as a key driver of the results. In computing the IQR, we extract the 25th and 75th percentiles of the individual densities under the assumption that probabilities are uniformly distributed within each interval.

5.2 Panel evidence

As previously discussed, the cross section analysis suffers from the shortcoming that it ignores any potential information associated with the variation of density forecast features over time. The linear panel regression in levels is intended to capture such effects. The results of the estimation of equation (8) are reported in Table 4a. In general, the findings from the panel regression are very consistent with the results of the cross-section regression described previously with $\beta_1 > 0$ and $\beta_2 < 0$ for each variable and each horizon. The finding of $\beta_2 < 0$, which represents the average impact of changes in variance within the full panel, again confirms the tendency for increases in variance to be associated with improved performance. Conversely, a decrease in variance is systematically associated with higher scores and worse performance. We would interpret the finding of a negative relation between density forecast variance and density score as providing further empirical evidence for the overconfidence of SPF forecasters highlighted previously. Once again the estimated coefficients suggest that this result is modestly stronger for GDP growth than it is for inflation.³⁰ The panel results also point to the positive relation between skewness and density score, an effect which – in contrast to the cross sectional regressions - tends to be significant in the full panel. This implies that more skewed distributions are associated with a worse performance, although this effect is quantitatively less important than either the mean forecast accuracy or the variance effects. The panel regressions also detect some significant contribution from fluctuations in tail risk to density performance. However, in the case of inflation, this is associated with higher scores (i.e. a worse performance). In contrast, for GDP growth at longer horizons (H=2) the estimated β_4 co-efficient is negative and is also significant. For this variable and at this horizon, this result implies that an increase in tail risk makes a contribution to lowering the density score and, thus, improving performance.

In considering the robustness of the above findings, it is possible to test the assumption of commonality of the estimated parameters across individuals. This is performed using the F-test discussed in Hsiao (2003) which tests the null hypothesis $H_0: \alpha_i = \alpha \beta_i = \beta$. Under the null, the Fisher distribution has $(K+1)(N-1)$ and $NT-(K+1)N$ degrees of freedom, where here K refers to the number of regressors used (in this case 4). Applying this procedure, the resulting test statistics are 0.0118 and 0.0218 for the case of the GDP models for H=1 and H=2 while for inflation the equivalent statistics are 0.0611 and

³⁰ The panel results using the IQR also confirm this systematic negative relationship. Additionally, the panel estimations were run over a shorter sample that excluded the most recent period influenced by the financial crisis (i.e. only with outcomes up to 2007Q4). This smaller sample also yields a systematic negative relation suggesting that our results are not driven by the crisis period alone.

0.1098 at H=1 and H=2. In no case are any of these statistics significant. This very much reflects the fact that the explanatory power of the regression with homogenous coefficients is already very high and as a result there is no significant incremental improvement when we expand the model to allow for heterogeneous parameters. Overall therefore this provides strong support for the pooled regressions that we estimate.

As discussed in Section 2, another potential source of distortion in making statistical inference is the presence of measurement error in the moments which are acting as explanatory variables in the panel regression. Following Grilliche and Hausmann (1986), comparing estimates of the panel regression in levels with an equivalent regression in first differences offers a relatively straightforward check that the results are not distorted by measurement error. The results of this regression are reported in Table 4b.³¹ Overall the regression in differences exhibits a very similar performance to the regression in levels, explaining over 80% of the fluctuations in performance in the case of inflation and more than 90% in the case of the GDP forecasts. Most importantly, we continue to observe the findings of $\beta_1 > 0$ and $\beta_2 < 0$ and these are statistically significant for each variable and each horizon. For the higher moments, some loss of significance in the estimated parameters is observed compared with the levels regression. Nonetheless, for longer horizon inflation densities (H=2), increased skewness continues to be associated with a significant deterioration in performance ($\beta_3 > 0$). Also, as in the levels panel regression, higher tail risk for the longer horizon GDP densities (H=2) continues to be associated with a better performance ($\beta_4 < 0$). Overall, this comparison points to considerable robustness in our findings particularly in relation to the mean forecast, the variance and skewness effects in the panel regressions.³² Our results on the information content of the higher moment features of the SPF densities, also contrasts somewhat with the evidence in Andrade, Ghysels and Idier (2012) whose work highlights a connection between the upper and lower quantiles of the inflation distribution and monetary policy and suggests some incremental explanatory power compared with the mean or consensus forecast. Their focus on monetary policy reactions, however, differs from our own focus on the performance of the predictive densities. Our results cast more doubt on the information content of the extremities of the SPF distributions in terms of prediction.

³¹ In the first difference regressions the F-statistics from the tests of the pooling of coefficients are 0.0220, 0.0425, 0.0549 and 0.0857 for GDP (H=1), GDP (H=2), Inflation (H=1) and Inflation (H=2) respectively. As was the case with the levels regression, these are never significant and thus tend to confirm the validity of the pooled model with homogenous parameters.

³² At the request of an anonymous referee, we have also conducted additional robustness checks by considering the inclusion of the IQR in the panel regression rather than the variance and obtained very similar results.

Table 5 reports the estimation results for equation (10) the panel regression in first difference and allowing for asymmetric effects. The results again tend to confirm the previous findings. In particular, the significant estimate of $\beta_2 < 0$ implies that positive changes in the variance of the predictive distributions tend to be associated with lower density scores or improved performance. Given that the estimation separates out positive from negative changes in the variance, the finding of $\beta_2 < 0$ could also be interpreted somewhat more positively in the sense that on average experts have been able to identify meaningful increases overtime in the “true” but unobserved uncertainty. In this respect, it is worth recalling the significant increase in variance observed for both GDP and inflation during the recent recession (2008-2009) and financial crisis (see also Figures 2 and 3). The estimated negative co-efficient on $\beta_3 < 0$ implies that stronger decreases in variance are systematically associated with higher scores. Comparing the estimates of β_2 and β_3 in Table 5, there is no evidence of significant asymmetry when considering separately the impact of increases and decreases in variance. Turning to the other higher moment features, the estimates for $\beta_4, \beta_5, \beta_6$ and β_7 in these asymmetric panel regressions are often insignificant, implying that whenever expert densities become more skewed or fatter tailed, this is invariably associated with no impact on forecast performance. An exception is the estimate of β_6 and β_7 for the two-year ahead GDP density, where higher tail probability mass is associated with an improvement in density performance, a result which matches the findings in Table 4a. In this case, the point estimates of β_6 and β_7 are almost identical pointing to no noticeable asymmetry when distinguishing instances of increases in tail risk from instances of decreases.

6. CONCLUSIONS

In this study we have examined using both cross section and panel regressions the possible link between density forecast performance and density features such as their location, spread, skewness and tail risk. Understanding this link is of relevance to both forecast users and forecast producers. In particular, the insights from our study are of interest to users of such surveys who may rely on specific density features when evaluating different policy options (e.g. relating to tail risks or possible skewness in the distributions). Moreover, density forecast producers, including those forecast producers responding to the SPF questionnaire, may be able to improve their forecast performance by gaining an understanding about how their density features have impacted on their historical density performance. Such a mode of analysis, responds to a clear need, as expressed in Leeper (2003) and more recently in Knüppel and Schulterfrankenfeld (2012)

to generate empirical evidence concerning the quality of information that is contained in density forecasts.

Our empirical evidence on the link between density forecast performance and density features relates to euro area GDP growth and inflation in the period since the launch of the single currency in 1999. A key strength of our study is the exploitation of micro data and the use of evaluation criteria which score the entire predictive densities. Controlling for the effects of common macroeconomic shocks, we apply cross-section and panel regression techniques to micro data from the euro area SPF. Our findings highlight the importance of mean forecast accuracy as a systematic determinant of individual density forecast performance: Experts with less accurate mean forecasts in general also have less accurate density forecasts. However, our empirical results suggest that many macroeconomic experts are operating in a way in which density performance could be systematically improved by correcting a downward bias in their variances. In other words, many experts responding to the SPF are underestimating uncertainty and could improve their density performance by simply increasing their variances. Aside from this shortcoming in second moment characteristics of the individual densities, other higher moment features, such as density skewness or kurtosis generally tend not to contribute significantly to enhancing individual density performance. Hence our results cast some doubt on the degree to which central banks should rely on these features of private sector density forecasts when evaluating macroeconomic risks. Increases in distributional skewness in absolute terms, for example, are invariably associated with a worse density performance – an effect which is statistically significant for both GDP growth and consumer price inflation. This result for private sector forecasts mirrors a similar finding in Knüppel and Schulterfrankenfeld (2012) in relation to central banks' own risk assessments.

Overall our analysis points to some scope for professional forecasters to potentially improve their density forecast performance, in particular by correcting upwards a noticeable downward bias in the variance of their densities. Moreover our results suggest that forecast users should exercise due caution when extracting information from higher moments of the SPF densities such as measures of skewness and, as in the recent study by Andrade, Ghysels and Idier (2012), tail risk. Looking forward, with the expansion of the SPF dataset over time, the scope to apply other density evaluation techniques will enlarge. In particular, it would be desirable to estimate and evaluate the probability integral transforms for individual SPF data for the euro area and to possibly extend the analysis conducted here to consider other density scoring rules.

References

Andrade, P., E. Ghysels and J. Idier, (2011), Tails of inflation forecasts and tales of monetary policy, Working Paper no. 407, Banque de France

Boero, G., J. Smith, and K. F. Wallis, (2012). The measurement and characteristics of professional forecasters' uncertainty, paper presented at The 32nd Annual International Symposium on Forecasting 24 - 27 June 2012, Boston, USA

Boero, G., J. Smith and K. F. Wallis (2011), Scoring rules and survey density forecasts, *International Journal of Forecasting*, 27(2), April-June, 379-393

Boero, G., J. Smith and K. F. Wallis (2008), Evaluating a three-dimensional panel of point forecasts: The Bank of England Survey of External Forecasters, *International Journal of Forecasting*, 24(2008): 354–367

Bowles, C., R. Friz, V. Genre, G. Kenny, A. Meyler and T. Rautanen (2010), An evaluation of the growth and unemployment rate forecasts in the ECB SPF, *Journal of Business Cycle Measurement and Analysis*, 2010(2):63-90

Casillas-Olvera, G. and D.A. Bessler (2006), Probability forecasting and central bank accountability, *Journal of Policy Modelling*, 28(2):223-234

Clements, M. P. (2012). Subjective and ex post forecast uncertainty, Warwick Economic Research Papers, No. 995, The University of Warwick.

Clements, M. P. (2010), Explanations of inconsistencies in survey respondent's forecasts, *European Economic Review*, 54(4):536-549

Clements, M. P. (2006), Evaluating the Survey of Professional Forecasters probability distributions of expected inflation based on the derived event probability forecasts, *Empirical Economics*, 31, 49-64

Conflitti C. (2011), Measuring Uncertainty and Disagreement in the European Survey of Professional Forecasters, *Journal of Business Cycle Measurement and Analysis*, 2011(2): 69-103

Corradi, V. and N. R. Swanson (2006), Predictive Density Evaluation, in: Handbook of Economic Forecasting (eds. Clive W.J. Granger, Graham Elliot and Allan Timmerman) Elsevier, Amsterdam: 197-284

Curtin, R. (2010), “Inflation expectations: theoretical models and empirical tests” in “Inflation Expectations” by P. Sinclair (ed.) *Routledge International Studies in Money and Banking*

D'Agostino, A., K. McQuinn and K. Whelan, (2012), Are Some Forecasters Really Better Than Others?, *Journal of Money Credit and Banking*, 44(4): 715-732

Davies, A. and K. Lahiri (1999), Re-examining the rational expectations hypothesis using panel data on multi-period forecasts, in C.Hsiao, M.H.Pesaran, K.Lahiri & L.F.Lee(Eds.), *Analysis of Panels and Limited Dependent Variable Models*, Cambridge: Cambridge University Press: 226–254

Davies, A. and K. Lahiri (1995), A new framework for analyzing survey forecasts using three-dimensional panel data, *Journal of Econometrics*, 68: 205–227

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

Diebold, F. X., Tay, A. S. and K. F. Wallis (1999), Evaluating density forecasts of inflation: the Survey of Professional Forecasters, in Engle R. and White H. (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, Oxford University Press, Oxford

Diebold, F. X., Gunther, T. and A. Tay (1998), Evaluating density forecasts with application to financial risk management, *International Economic Review*, 39(4):863–883

ECB (2014a), Fifteen years of the ECB Survey of Professional Forecasters, ECB Monthly Bulletin, January 2014

ECB (2014b), Results of a second special questionnaire for participants in the ECB Survey of Professional Forecasters, January 2014. Available at <http://www.ecb.europa.eu/stats/prices/indic/forecast/html/index.en.html>

Engelberg, J., C.F. Manski, and J. Williams (2009), Comparing the point predictions and subjective probability distributions of professional forecasters, *Journal of Business and Economic Statistics*, 27(1):30-41

Epstein, E.S. (1969), A scoring system for probability forecasts of ranked categories, *Journal of Applied Meteorology*, 8:985-987

Garrat, T., J. Mitchell and S. Vahey (2013), Probability Forecasting for inflation warnings from the US Federal Reserve, paper presented at the 9th Annual joint Cirano-Cireq Workshop on data revisions in economic forecasting and policy, Montreal, 10-11 October 2013

Genre, V., G. Kenny, A. Meyler and A. Timmermann (2013), Combining expert forecasts: Can anything beat the simple average?, *International Journal of Forecasting*, 29(1): 108-121

Giordani, P., and P. Söderlind (2006), Is there evidence of pessimism and doubt in subjective distributions? Implications for the equity premium puzzle, *Journal of Economic Dynamics and Control*, 30(6):1027-1043

Giordani, P., and P. Söderlind (2003), Inflation Forecast Uncertainty, *European Economic Review*, 47(6): 1037–1059.

Gneiting, T. and A. E Raftery (2007), Strictly Proper Scoring Rules: Prediction and Estimation, *Journal of the American Statistical Association*, March 2007, 102(477):,359-378

Grilliche Z. and J. Hausmann (1986), Errors in variables in panel data, *Journal of Econometrics*, 1986(31): 93-118

Hsiao C. (2003), Ananalysis of Panel Data, 2nd Edition, *Cambridge University Press*

Kahneman, D and A. Tversky (2000), Choices, Values and Frames, *Cambridge University Press*, Russel Sage Foundation

Keane, M. P. and D. E. Runkle (1990), Testing the rationality of price forecasts: new evidence from panel data, *American Economic Review*, 80(4), September 1990:714-735

Kenny, G., T. Kostka and F. Masera (2014), How informative are the subjective density forecasts of Macroeconomists?, forthcoming, *Journal of Forecasting*, also available as ECB Working Paper No. 1446

Kenny, G., T. Kostka and F. Masera (2013), Can macroeconomists forecast risk? Event-based evidence from the ECB SPF, ECB Working Paper No. 1540

Knüppel, M. and G. Schulerfrankenfeld (2012), How informative are central bank assessments of macroeconomic risks, *International Journal of Central Banking*, 8(3):87-139

Lahiri, K. and J. G. Wang (2013), Evaluating probability forecasts for GDP declines using alternative methodologies, *International Journal of Forecasting* 29 (2013) 175–190

Lahiri, K. and J. G. Wang (2007), The value of probability forecast as predictors of economic downturns, *Applied Economic Letters*, 14(14), 11-14

Lahiri, K., C. Teigland and M. Zaporowski, (1988), 'Interest rates and the subjective probability distribution of inflation forecasts', *Journal of Money, Credit, and Banking*, 20(2):233–248

Leeper, E. M. (2003), An Inflation reports Report, *NBER Working Paper* No. 10089, November 2003

Mitchell, J. and K. Wallis (2011), Evaluating density forecasts: Forecast combinations and model mixtures, calibration and sharpness, *Journal of Applied Econometrics*, 26(6), 1023-1040

Peseran, M. H. (2006), Estimation and inference in heterogenous panels with a multifactor error structure, *Econometrica*, 74(4), 967-1012

Rich, R., Song, J. and J. Tracey (2012), The measurement and behaviour of uncertainty: Evidence from the ECB Survey of Professional Forecasters, Federal Reserve bank of New York, *Staff Report* No. 588

Rich, R. and J. Tracy (2010), The relationship among expected inflation, disagreement, and uncertainty: Evidence from matched point and density forecasts, *The Review of Economics and Statistics*, 92(1), 200-207

Sims, Christopher A. (2003): "Implications of Rational Inattention", *Journal of Monetary Economics*, 50(3): 665-690

Tay, A.S. and Wallis, K.F. (2000), Density forecasting: a survey, *Journal of Forecasting*, 19:235-254. Reprinted in *A Companion to Economic Forecasting* (M.P. Clements and D.F. Hendry, eds.), 45-68. Oxford: Blackwell, 2002

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, 1st edition, *The MIT Press*

Zarnowitz, V. and L. A. Lambros, (1987), Consensus and uncertainty in economic prediction, *Journal of Political Economy*, 95(3), 591-621

Figure 1: Ranked Probability Scores for SPF density forecasts for GDP growth and HICP Inflation

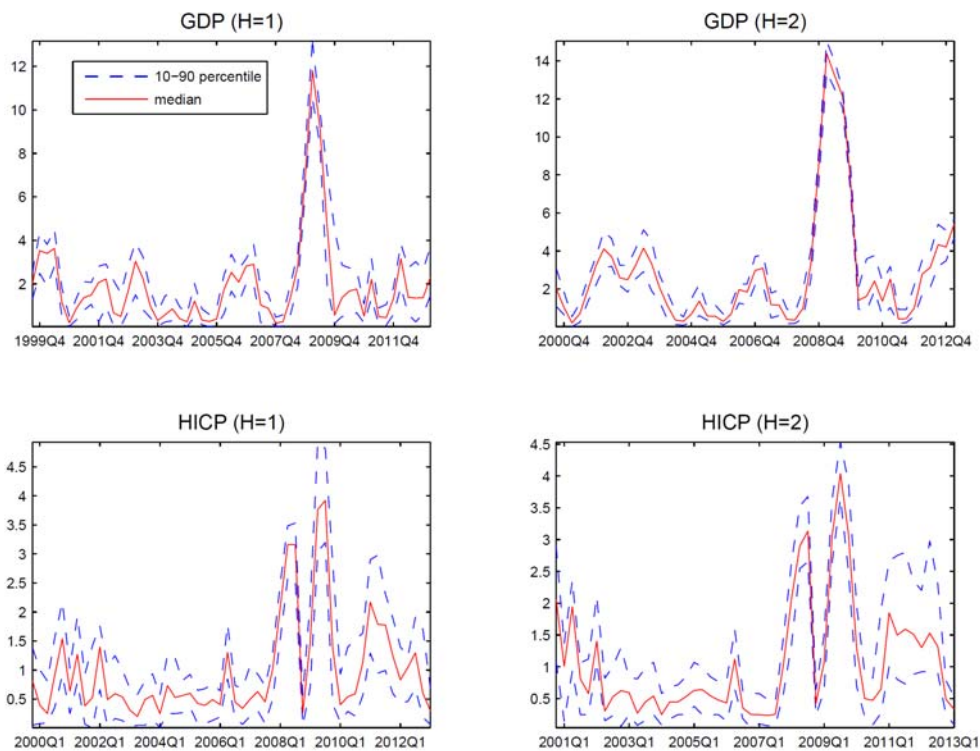


Figure 2: SPF density moments for GDP growth (H=1)

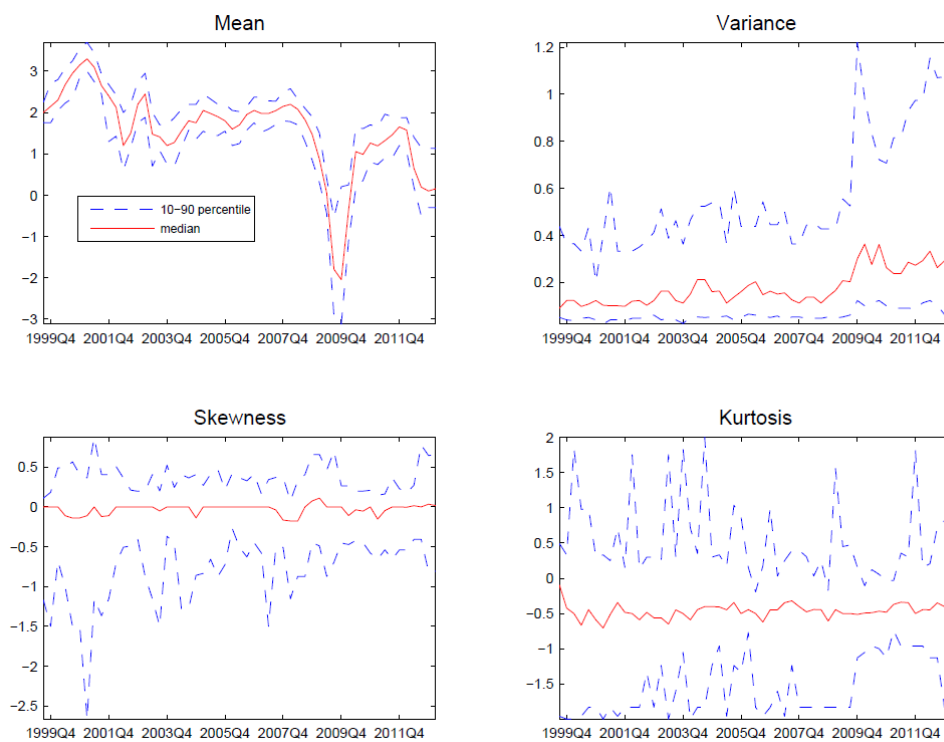


Figure 3: SPF density moments for HICP Inflation (H=1)

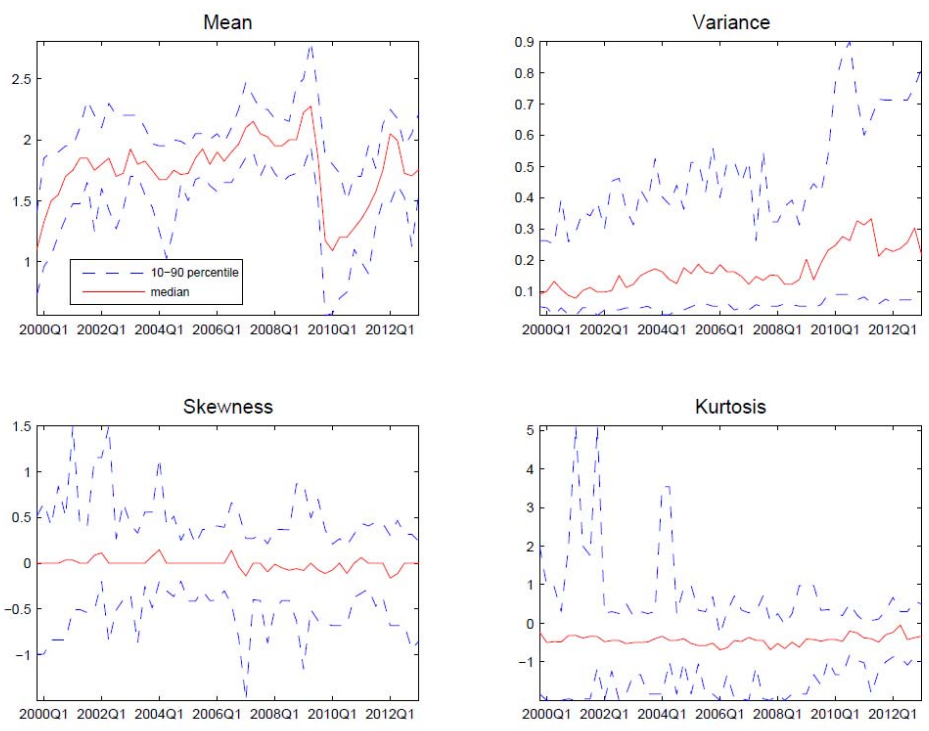


Table 1: Sample statistics: SPF forecasts of GDP Growth
(Sample Averages, 1999Q1-2013Q1)

	μ	ME	MAE	σ^2	γ	ζ
				<i>H=1</i>		
min	1.36	0.08	1.06	0.09	0.10	-0.77
10%	1.40	0.12	1.09	0.10	0.16	-0.76
25%	1.51	0.23	1.11	0.12	0.20	-0.58
50%	1.58	0.30	1.20	0.20	0.24	-0.39
75%	1.66	0.37	1.23	0.35	0.33	-0.26
90%	1.75	0.46	1.26	0.44	0.45	0.03
max	1.81	0.53	1.29	0.79	0.63	0.54
				<i>H=2</i>		
min	1.86	0.79	1.53	0.11	0.07	-0.97
10%	1.91	0.84	1.55	0.13	0.17	-0.82
25%	1.99	0.91	1.58	0.16	0.19	-0.65
50%	2.08	1.01	1.63	0.23	0.26	-0.41
75%	2.11	1.04	1.65	0.37	0.32	-0.28
90%	2.16	1.09	1.69	0.56	0.40	-0.24
max	2.20	1.12	1.72	1.06	0.53	0.77

Note: ME denotes the average forecast error, where the error is defined as the mean forecast minus the outcome, MAE denotes the mean absolute error, σ^2 denotes the variance, γ the absolute skewness and ζ the excess kurtosis.

Table 2: Sample statistics: SPF forecasts of Inflation
(Sample Averages, 1999Q1-2013Q1)

	μ	ME	MAE	σ^2	γ	ζ
				<i>H=1</i>		
min	1.65	-0.47	0.55	0.09	0.14	-0.94
10%	1.66	-0.46	0.57	0.09	0.16	-0.85
25%	1.70	-0.42	0.64	0.12	0.20	-0.56
50%	1.75	-0.37	0.67	0.20	0.25	-0.34
75%	1.80	-0.32	0.70	0.28	0.34	-0.20
90%	1.90	-0.22	0.72	0.36	0.48	0.22
max	1.99	-0.14	0.75	0.61	0.78	1.26
				<i>H=2</i>		
min	1.59	-0.56	0.58	0.09	0.08	-1.07
10%	1.71	-0.44	0.61	0.11	0.13	-0.65
25%	1.74	-0.40	0.66	0.16	0.17	-0.59
50%	1.82	-0.32	0.71	0.23	0.23	-0.40
75%	1.85	-0.29	0.74	0.33	0.26	-0.27
90%	1.97	-0.18	0.76	0.51	0.38	0.13
max	2.12	-0.02	0.87	0.76	0.60	0.97

Note: ME denotes the average forecast error, where the error is defined as the mean forecast minus the outcome, MAE denotes the mean absolute error, σ^2 denotes the variance, γ the absolute skewness and ζ the excess kurtosis.

Table 3a: Distribution features and density forecast performance: Cross Sectional Evidence

$$L_i = \alpha + \beta_1 |y - \mu_i| + \beta_2 \sigma_i^2 + \beta_3 \gamma_i + \beta_4 \zeta_i + \varepsilon_i$$

	α	β_1	β_2	β_3	β_4	N	R ²
GDP (H=1)	-0.26* (0.15)	1.97** (0.11)	-0.44** (0.05)	0.29** (0.13)	-0.02 (0.05)	24	0.96
GDP (H=2)	0.11 (0.15)	1.73** (0.09)	-0.45** (0.05)	0.33 (0.21)	0.00 (0.05)	25	0.95
Inflation (H=1)	-0.02 (0.1)	1.61** (0.14)	-0.22** (0.07)	0.14 (0.14)	-0.01 (0.03)	24	0.90
Inflation (H=2)	-0.05 (0.1)	1.64** (0.12)	-0.15** (0.06)	0.11 (0.15)	-0.01 (0.03)	25	0.90

Table 3b: Distribution features and density forecast performance: Cross Sectional Evidence

$$L_i = \alpha + \beta_1 |y - \mu_i| + \beta_2 IQR_i + \beta_3 \gamma_i + \beta_4 \zeta_i + \varepsilon_i$$

	α	β_1	β_2	β_3	β_4	N	R ²
GDP (H=1)	0.01 (0.12)	1.92** (0.09)	-0.45** (0.04)	0.15 (0.09)	-0.03 (0.03)	24	0.97
GDP (H=2)	1.04** (0.27)	1.39** (0.17)	-0.62** (0.05)	-0.02 (0.14)	-0.03 (0.04)	25	0.92
Inflation (H=1)	0.22* (0.1)	1.39** (0.14)	-0.21** (0.05)	0.12 (0.09)	0.00 (0.02)	24	0.90
Inflation (H=2)	0.10 (0.14)	1.5** (0.16)	-0.13* (0.06)	0.10 (0.15)	-0.02 (0.03)	25	0.84

Note: Cross sectional regression using OLS where all variables are expressed as sample averages over the period 1999Q1-2013Q1. Coefficients significant at the 5% level are indicated with "***" and at the 10% level with "*". The standard deviation of the coefficient estimate are reported in parenthesis and rounded to two decimal points.

Table 4a: Distribution features and density forecast performance: Panel regression in levels

$$L_{i,t+\tau} = \alpha + \beta_0 D_t + \beta_1 |y_{i,t+\tau} - \mu_{i,t+\tau}| + \beta_2 \sigma_{i,t+\tau}^2 + \beta_3 \gamma_{i,t+\tau} + \beta_4 \zeta_{i,t+\tau} + \varepsilon_{i,t+\tau}$$

	α	β_1	β_2	β_3	β_4	N	R ²
GDP (H=1)	0.16** (0.05)	1.76** (0.02)	-0.42** (0.02)	0.03** (0.02)	0.00 (0.01)	1320	0.99
GDP (H=2)	1.33** (0.07)	1.47** (0.02)	-0.53** (0.02)	0.05** (0.02)	-0.03** (0.01)	1275	0.99
Inflation (H=1)	0.14** (0.03)	1.59** (0.02)	-0.33** (0.03)	0.02* (0.01)	0.01* (0.01)	1296	0.97
Inflation (H=2)	0.13** (0.04)	1.49** (0.02)	-0.12** (0.03)	0.03* (0.02)	0.01** (0.01)	1250	0.96

Table 4b: Distribution features and density forecast performance: Panel regression in first differences

$$\Delta L_{i,t+\tau} = \alpha + \beta_0 D_t + \beta_1 \Delta |y_{i,t+\tau} - \mu_{i,t+\tau}| + \beta_2 \Delta \sigma_{i,t+\tau}^2 + \beta_3 \Delta \gamma_{i,t+\tau} + \beta_4 \Delta \zeta_{i,t+\tau} + \varepsilon_{i,t+\tau}$$

	α	β_1	β_2	β_3	β_4	N	R ²
GDP (H=1)	0.00 (0.01)	1.99** (0.00)	-0.55** (0.04)	0.01 (0.02)	0.00 (0.01)	1296	0.92
GDP (H=2)	0.01** (0.00)	1.93** (0.01)	-0.80** (0.04)	0.01 (0.03)	-0.05** (0.01)	1250	0.92
Inflation (H=1)	0.00** (0.00)	1.57** (0.02)	-0.45** (0.1)	-0.03 (0.03)	0.01 (0.01)	1272	0.85
Inflation (H=2)	0.00** (0.00)	1.58** (0.02)	-0.17** (0.08)	0.06* (0.03)	0.01 (0.01)	1225	0.82

Note: Robust parameter estimates using feasible GLS are reported, estimated over the sample 1999Q1-2013Q1 and including a correction for serial and cross-sectional correlation. Coefficients significant at the 5% level are indicated with "***" and at the 10% level with "*". The standard deviation of the coefficient estimate is reported in parenthesis and rounded to two decimal points.

Table 5: Distributional features and density forecast performance: Panel regression in 1st differences with asymmetries

$$\Delta L_{i,t+\tau} = \alpha + \beta_0 D_t + \beta_1 \Delta |y_{t+\tau} - \mu_{i,t+\tau}| + \beta_2 (\Delta \sigma_{i,t+\tau}^2)^+ + \beta_3 (\Delta \sigma_{i,t+\tau}^2)^- + \beta_4 \Delta \gamma_{i,t+\tau}^+ + \beta_5 \Delta \gamma_{i,t+\tau}^- + \beta_6 \Delta \zeta_{i,t+\tau}^+ + \beta_7 \Delta \zeta_{i,t+\tau}^- + \varepsilon_{i,t+\tau}$$

	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	N*T	R ²
	<i>Δ Ranked Probability Score</i>									
GDP (H=1)	0.53** (0.05)	1.73** (0.02)	-0.47** (0.05)	-0.49** (0.05)	0.00 (0.02)	0.00 (0.02)	0.01 (0.01)	0.01 (0.01)	1296	0.98
GDP (H=2)	0.66** (0.05)	1.47** (0.02)	-0.52** (0.03)	-0.58** (0.03)	0.03 (0.03)	0.04 (0.03)	-0.03** (0.01)	-0.03** (0.01)	1250	0.98
Inflation (H=1)	0.27** (0.04)	1.59** (0.02)	-0.43** (0.06)	-0.4** (0.06)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	1272	0.96
Inflation (H=2)	0.24** (0.04)	1.47** (0.02)	-0.09* (0.05)	-0.15** (0.05)	0.03 (0.02)	0.02 (0.02)	0.01 (0.01)	0.01* (0.01)	1225	0.95

Note: Parameter estimates using feasible GLS are reported, estimated over the period 1999Q1-2013Q1 and including a correction for serial correlation. Coefficients, significant at the 5% level are indicated with "***" and at the 10% level with "*". The standard deviation of the coefficient estimate is reported in parenthesis and rounded to two decimal points.