NATIONAL BANK OF POLAND WORKING PAPER No. 87

Predictivistic Bayesian Forecasting System

Andrzej Kocięcki, Marcin Kolasa, Michał Rubaszek

Warsaw 2011

Andrzej Kocięcki - National Bank of Poland. E-mail: andrzej.kociecki@nbp.pl.

Marcin Kolasa - National Bank of Poland, Warsaw School of Economics. E-mail: marcin.kolasa@nbp.pl.

Michał Rubaszek - National Bank of Poland, Warsaw School of Economics. E-mail: michal.rubaszek@nbp.pl.

We thank M. Jarociński, R. Kokoszczyński, H. Kowalczyk, B. Mazur, A. Pajor, M. Pipień for useful comments to the previous versions of this article. The views expressed in this paper are those of the authors and not necessarily those of the National Bank of Poland.

Design:

Oliwka s.c.

Layout and print:

NBP Printshop

Published by:

National Bank of Poland
Education and Publishing Department
00-919 Warszawa, 11/21 Świętokrzyska Street
phone: +48 22 653 23 35, fax +48 22 653 13 21

http://www.nbp.pl

Contents

List of Tables

List of Figures

Abstract

This paper introduces a formal method of combining expert and model density forecasts when the sample of past forecasts is unavailable. It works directly with the expert forecast density and endogenously delivers weights for forecast combination, relying on probability rules only. In the empirical part, we illustrate how our framework can be applied in forecasting US inflation by mixing density forecasts from an autoregressive model and the Survey of Professional Forecasters.

Keywords: Combining density forecasts; Forecast evaluation; Bayesian inference; Predictivism

JEL classification: C11, C53

Non-technical summary

The topic of combining forecasts has received considerable attention both in the theoretical literature and empirical applications. This can be explained by a well-documented finding that merging point forecasts can improve forecast accuracy. In most cases, a typical combined point forecast is calculated as a weighted average of individual forecasts, where weights can be fixed and computed using classical or Bayesian methods, or can be time-varying. The literature on density forecast combination is much more sparse. As a result, a simple average of cumulative distribution functions is used in most empirical applications.

Another important divergence between the literature and practical needs is that the main focus of the former has evolved into developing effective methods of combining forecasts from various estimated models. At the same time, one of the main problems faced by decision makers is how to incorporate expert judgment into model-based density forecasts. For example, most central banks have developed sophisticated macroeconomic models that are used for forecasting, but due to the lack of formal methods resort to ad hoc techniques while adjusting their model forecasts. This type of dilemmas is characteristic for many other institutions and firms, which in most cases use purely judgmental forecasts, apply some simple expert corrections or combine forecasts with ad hoc weights, for example assuming them equal for an expert and a model.

At the same time, potential gains from merging expert and model-based density forecasts can be quite substantial as they seem to be complementary at least in some respects. In particular, there are reasons to expect that combined forecasts could benefit from point accuracy of experts and more valid confidence intervals generated by econometric models.

This paper contributes to the literature by introducing a new formal method of combining expert and model-based density forecasts. Since the method applies Bayesian inference and comes close in the essence to the predictivism, we dub it as the Predictivistic Bayesian Forecasting System (PBFS). A crucial feature of our approach is that the weights it delivers are endogenously determined within the framework and rely on probability rules only, which contrasts with the literature surveyed above. Its output can be interpreted as a result of merging information from three sources: an expert, a statistical model and the economic theory.

Importantly, in comparison to the standard methods of combining density forecasts, our approach can be applied even if the previous history of forecasts is unavailable, which is often the case for expert-based forecasts. Another important advantage of the PBFS is that it is able to work directly with a density forecast of the expert. This is in contrast to the standard Bayesian approach in incorporating expert knowledge into statistical models, which hinges on the information conversion from the expert's density forecast into the prior for statistical model parameters.

In the empirical part of the article we demonstrate how to implement the PBFS algorithm to combine US inflation forecasts from a simple autoregressive model and the Survey of Professional Forecasters (SPF). Given simplicity of our application, the results of the out-of-sample forecasting performance evaluation look very promising. Even though point forecasts from the PBFS are less accurate than those from experts alone, our framework clearly outperforms the SPF in terms of interval forecasts.

1

1 Introduction

The topic of combining forecasts has received considerable attention both in the theoretical literature and empirical applications. This can be explained by a well-documented finding that merging point forecasts can improve forecast accuracy (Timmermann, 2006). In most cases, a typical combined point forecast is calculated as a weighted average of individual forecasts, where weights can be fixed and computed using classical (Elliott and Timmermann, 2004) or Bayesian methods (Min and Zellner, 1993), or can be time-varying as e.g. in Guidolin and Timmermann (2009).

The literature on density forecast combination is much more sparse. As a result, a simple average of cumulative distribution functions is used in most empirical applications. Only recently some authors have proposed methods that can be used to combine density forecasts (e.g. Hall and Mitchell, 2007), but this research area remains still relatively unexplored. This might be seen as surprising as density forecasts are widely used in various domains like macroeconomic forecasting, risk analysis or weather forecasting. Clearly, demand for methods of combining density forecasts is far from satiated.

Another important divergence between the literature and practical needs is that the main focus of the former has evolved into developing effective methods of combining forecasts from various estimated models. At the same time, one of the main problems faced by decision makers is how to incorporate expert judgment into model-based density forecasts. For example, most central banks have developed sophisticated macroeconomic models that are used for forecasting, but due to the lack of formal methods resort to ad hoc techniques while adjusting their model forecasts (Pagan, 2003). This type of dilemmas is characteristic for many other institutions and firms, which in most cases use purely judgmental forecasts or apply some simple expert corrections (Lawrence et al., 2006).

It has to be stressed that merging expert judgment with output from statistical models using the existing tools may be highly problematic, even for point forecasts. Since in many relevant cases the sample of past expert forecast errors is short or unavailable, standard methods of obtaining optimal weights cannot be applied. As a result, the only option for decision makers is to combine forecasts with ad hoc weights, for example assuming them equal for an expert and a model.

At the same time, potential gains from merging expert and model-based density forecasts can be quite substantial as they seem to be complementary at least in some respects. The empirical literature suggests that the accuracy of expert-based point forecasts for standard macroeconomic variables tend to be higher than those obtained from state-of-art estimated models (Clark and McCracken, 2006; Rubaszek and Skrzypczynski, 2008). On the other hand, it is well established that experts tend to significantly underestimate uncertainty.[1] As demonstrated by Giordani and Soderlind (2003), this is also evident for individual inflation forecasts in the Survey of Professional Forecasters (SPF). All in all, there are reasons to expect that combined forecasts could benefit from point accuracy of experts and more valid confidence intervals generated by econometric models.

This paper contributes to the literature by introducing a new formal method of combining expert and model-based density forecasts. Since the method applies Bayesian inference and comes close in the essence to the predictivism, we dub it as the Predictivistic Bayesian Forecasting System (PBFS). A crucial feature of our approach is that the weights it delivers are endogenously determined within the framework and rely on probability rules only, which contrasts with the literature surveyed above. Moreover, in comparison to the standard methods of combining density forecasts (e.g. using log score-based averaging, see Hall and Mitchell, 2007; Jore et al., 2010), our approach can be applied even if the previous history of forecasts is unavailable, which is often the case for expert-based forecasts. Another important advantage of the PBFS is that it is able to work directly with a density forecast of the expert. This is in contrast to the standard Bayesian approach in incorporating expert knowledge into statistical models, discussed e.g. in Kadane (1980), Kadane et al. (1980) or Garthwaite et al. (2005), which hinges on the information conversion from the expert's density forecast into the prior for statistical model parameters. As a result, the PBFS can absorb any, even highly nonstandard form of the expert's forecast probability density function (pdf), while in the standard Bayesian approach the expert's pdf is restricted by the form of the likelihood and the prior assumed by the modeller.

In the empirical part of the article we demonstrate how to implement the PBFS

[1]See Lawrence et al. (2006) for a discussion of possible reasons.

algorithm to combine US inflation forecasts from a simple autoregressive model and the SPF. Given simplicity of our application, the results of the out-of-sample forecasting performance evaluation look very promising. Even though point forecasts from the PBFS are less accurate than those from experts alone, our framework clearly outperforms the SPF in terms of interval forecasts.

The rest of this paper is structured as follows. Section 2 lays out the PBFS framework. The algorithm for a simple autoregressive model and an expert forecast density of a known type is presented in section 3. Section 4 applies this example to forecasting US inflation. Section 5 discusses the out-of-sample forecasting performance of the PBFS and its individual components. The last section concludes.

2 The PBFS setup

Our departure point is the following factorization of the joint past and future data pdf:

$$p(y^*|y) = \frac{p(y|y^*)p(y^*)}{p(y)} \tag{1}$$

where y and y^* denote past and future realization of some vector of variables, respectively, and $p(y) = \int p(y|y^*)p(y^*)dy^*$.

The left-hand side of equation (1) is the posterior forecast pdf, which is a result of updating some prior beliefs about future observations $p(y^*)$, after having confronted them with the past data. In the PBFS, we will assume that the prior is supplied by an expert (E), while $p(y|y^*)$ comes from a statistician (S). Hence, equation (1) can be seen as a formula for merging expert and model-based knowledge into a combined posterior forecast (C). We write this explicitly as:

$$p_C(y^*|y) = \frac{p_S(y|y^*)p_E(y^*)}{p_C(y)} \tag{2}$$

where $p_C(y) = \int p_S(y|y^*)p_E(y^*)dy^*$.

Before we proceed with further derivations, two comments are in order. First, one can note that the way we write formula (1) abstracts away from any model parameters as all densities are functions of observables only. This is in contrast to the standard Bayesian inference, where the departure point is the model $p(y, \theta)$, with θ standing for a vector of parameters. As we will see, the parameters do eventually appear in the PBFS once we start characterizing the statistician's input. However, they are introduced indirectly and only for the ease of modeling. Therefore, our approach is in line with the inferential framework advocated by predictivists, exemplified by Geisser (1993), Bernardo and Smith (1994, pp. 165-167), Lane (1985) and, according to Stigler (1982), Thomas Bayes himself. We note that many Bayesian thinkers support the idea that probability statements should be confined to observables (see de Finetti, 1975, pp. 200-201; Lindley, 2006, p. 77).

Second, some may question describing expert judgment by $p_E(y^*)$ on the grounds that experts usually take into account the past data while formulating their forecasts.

This is an analogous problem to the one that bothered Thomas Sargent (Klamer, 1984, p. 75):

> *It's critical that you didn't look at the data before. Acquiring priors from the data and then going back and using the same data does not seem to be right. Such objections are important for me. I think about them all the time but I don't let them stop me.*

In general, the "data double-use" problem has already been extensively discussed by many authors (see e.g. Poirier, 1995, and references therein). We note, however, that the idea of eliciting the prior after seeing the data has been already promoted in the Bayesian literature by Hill (1985, 1992) and underlies much of applied Bayesian econometric research, take the estimation of dynamic stochastic general equilibrium models for instance (Smets and Wouters, 2003). In fact, there are fundamental reasons why the view that the prior could be formed after seeing the data is legitimate. The argumentation is as follows. First, as it was justified by Good (1950, p. 41), Jeffreys (1961, chap. 1) or de Finetti (1972, p. 150), the Bayes formula should be regarded as a kind of inductive tool. Second, inductive inference is correct if it does not miss any information available about the inferential object (de Finetti, 1975, p. 196 and Fisher, 1956, p. 55).

In our model the inferential object is $p(y^*, y)$, which was decomposed (but only for our convenience) as $p(y|y^*)p(y^*)$. Note that in most theories of probability the formula of conditional probability is just a definition. Since we do not make "inductive" mistake only if we use all available information while reasoning about $p(y^*, y)$, the same is true for $p(y|y^*)$ and $p(y^*)$ taken individually. We come to a conclusion that the essential defining property of $p(y^*)$ is not that it is *a priori* but the marginal density. Provided that $p(y^*)$ does not depend functionally on y our inference is fully coherent, i.e. we obey the probability rules. There are no other constraints imposed on the expert. This is perfectly consistent with the view about the induction contained in Keynes (1921, pp. 305-306). In his pre-statistical language, $p(y|y^*)p(y^*)$ is called a hypothesis. He states that *the question as to whether a particular hypothesis happens to be propounded before or after [observations] examination is quite irrelevant* because *it is the union of prior knowledge, with the inductive grounds which arise out of the immediate instances, that lends weight to an hypothesis, and not the occasion on which the hypothesis is first* *proposed*. The irrelevance of the moment when we elicit the marginal density, which is better known as a priori density, was also emphasized by Hill (1992) and de Finetti (1975, p. 198). Lastly, we note that similar defence of our position may be based on Carnap (1950, pp. 211-212, 253 and 494). For the above reasons, we claim that the use of expert judgment as a prior in our framework cannot be dismissed as illegitimate.

Now we can move to defining the statistician's part. If she is a Bayesian, the relevant component of formula (2) can be written as:

$$p_S(y|y^*) = \frac{p_S(y^*|y)p_S(y)}{p_S(y^*)} \tag{3}$$

The statistician uses an econometric model, so all terms in (3) are marginal densities, with the model parameters θ integrated out. In particular:

$$p_S(y^*|y) = \int p_S(y^*|y,\theta)p_S(\theta|y)d\theta \tag{4}$$

is the standard model-based Bayesian forecast,

$$p_S(y) = \int p_S(y|\theta)p_S(\theta)d\theta \tag{5}$$

is the marginal data density implied by the model and

$$p_S(y^*) = \int p_S(y^*|y)p_S(y)dy = \int p_S(y^*|\theta)p_S(\theta)d\theta \tag{6}$$

can be called a prior forecast density of the statistician as it summarizes all model-based information on the future values of variables of interest, without a direct use of knowledge about their past realizations. We note that $p_S(y^*|\theta)$ for stationary models is the unconditional distribution of observations.

After inserting (3) into (2) and absorbing into the proportionality sign all densities that do not depend on y^*, we obtain the operational version of the PBFS:

$$p_C(y^*|y) \propto \frac{p_S(y^*|y)}{p_S(y^*)}p_E(y^*) \tag{7}$$

Formula (7) shows how to combine forecast densities from an econometric model with those generated by experts. Importantly, the implicit weights attached to these two sources of knowledge are endogenously determined within the PBFS framework and rely on probability rules only. This is in contrast to most of the literature, resorting to ad hoc weights (e.g. finite mixture distribution suggested by Wallis, 2005) or relying on the decision maker's assessment of the credibility of an expert (Morris, 1974).

The model-based input to the operational version of the PBFS consists of two parts. One, defined in equation (4), is just a standard Bayesian forecast pdf. However, in the PBFS the statistician is also required to provide the same object as the expert, i.e. the marginal pdf for future observations, given by equation (6). As a result, the PBFS delivers a framework for formal use of three sources of information in the forecasting process:

- expert judgment $p_E(y^*)$,
- conditional forecast from a statistical model $p_S(y^*|y)$,
- marginal density $p_S(y^*)$.

The last term can incorporate knowledge on the long-run dynamics implied by the economic theory. To see this, note that according to equation (6) $p_S(y^*)$ is a weighted average of the unconditional density for observations with respect to the prior of θ. Assuming stationarity, we can calibrate this prior so as it induces reasonable long-run dynamics for the economic variables of interest, as proposed e.g. by Villani (2009) for vector autoregressions.

3 PBFS algorithm for AR(1) model

The main difficulty with using the PBFS is that its output, i.e. the combined posterior forecast pdf, cannot be derived analytically, except for some uninteresting degenerate cases. This is very often true even if all densities appearing on the right-hand side of equation (7) are obtainable in an analytical form. Therefore, sampling methods have to be applied.

In order to illustrate how the PBFS can be used in practice, we consider its very simple, yet unobtainable analytically variant. Let us assume that the forecast is formulated for one period ahead and the statistician's part is based on the stationary AR(1) model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \tag{8}$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and $|\beta_1| < 1$. The statistician is a Bayesian and makes the following assumptions on the prior distributions of the model's parameters:

$$\begin{aligned} \beta|\sigma^2 &\sim N_{tr}(\underline{\beta}, \sigma^2 \mathbf{I}) \\ \sigma^2 &\sim IG(\underline{v}_1, \underline{v}_2) \end{aligned} \tag{9}$$

where $\beta = (\beta_0, \beta_1)$, the underlined symbols are hyperparameters, IG is the inverse gamma distribution and N_{tr} denotes the truncated normal distribution over the interval consistent with stationarity, i.e. when $|\beta_1| < 1$. As regards the expert, we assume that her forecast is given by a pdf of a known type.

In the simple illustration presented in this section $p_E(y^*)$ has an analytical form. The other components of (7), i.e. those attributed to the statistician, cannot be obtained analytically. This is due to the stationarity restriction $|\beta_1| < 1$ for (4) and the fact that in (6) the parameters enter $p_s(y^*|\theta)$ through unconditional moments $E(y^*) = \frac{\beta_0}{1-\beta_1}$ and $var(y^*) = \frac{\sigma^2}{1-\beta_1^2}$.

Our strategy to compute (4) and (6) is as follows. To obtain (4), we draw N times from the joint posterior of θ and approximate the value of $p_S(y^*|y)$ at y^* by $\frac{1}{N}\sum_{i=1}^{N} p_S(y^*|y, \theta^{(i)})$, where $\theta^{(i)}$ denotes the i-th draw from the posterior. In the case of

(6), we exploit the fact that analytical integration with respect to β_0 and σ^2 is possible. Then, drawing from the marginal prior for β_1, which is t-Student, we approximate the value of $p_S(y^*)$ at y^* as $\frac{1}{N}\sum_{i=1}^{N} p_S(y^*|\beta_1^{(i)})$. We note that $p_S(y^*|\beta_1^{(i)})$ is also t-Student.[2]

Having found a way to approximate $p_S(y^*|y)$ and $p_S(y^*)$, standard Bayesian sampling techniques can be used to obtain the final PBFS output, i.e. $p_C(y^*|y)$ defined in equation (7). In our application presented in the next section we will rely on the random walk Metropolis-Hastings (MH) algorithm. It has to be noted that even though the numerical approximation of the integrals discussed above is conceptually simple, the need to evaluate them in every step of the MH chain will significantly slow down the algorithm compared to its simple applications. Hence, in general, combining forecasts using the PBFS can be relatively time-consuming.

[2]Detailed derivations are available from the authors upon request.

4 An application to US inflation

We illustrate the typical output one can get from the PBFS by combining interval forecasts of US annual GDP deflator inflation from the SPF with the AR(1) model. We skip a detailed discussion of the SPF and its interval inflation forecasts, referring an interested reader to exhaustive descriptions in Croushore (1993), Diebold et al. (1999) or Giordani and Soderlind (2003). For the purpose of our study it is only important to mention that each forecaster participating in the survey is asked to attach a certain probability to a number of bins in which annual inflation might fall in the current and next year. The answers are published in the form of interval forecasts.

To construct the PBFS forecasts we proceed as follows:

1. We take the interval inflation forecast from the SPF for a given year τ, formulated in the first quarter of the same year. Following the arguments in Giordani and Soderlind (2003), we use the probabilities attached to each of the ten bins to approximate the expert density forecast for year τ, i.e. $p_{E,\tau}(y^*)$, with a Gaussian distribution, assuming that all probability mass are located at the interval midpoints.

2. We set the priors for the AR(1) at $\underline{\beta}_0 = 1$, $\underline{\beta}_1 = 0.5$, $\underline{v}_1 = 2$ and $\underline{v}_2 = 1$.

3. We use annual inflation data, taken from the real-time data set for macroeconomists (Croushore and Stark, 2001), to evaluate $p_{S,\tau}(y^*|y)$ and $p_{S,\tau}(y^*)$, where y denotes real-time data available in the first quarter of year τ, i.e. inflation time series from 1948 to year $\tau - 1$. Integrals given in formulas (4) and (6) are calculated as explained in section 3 with N=2,500 draws.

4. We apply the MH algorithm with 100,000 iterations, of which 10,000 initial draws are burned, to obtain the PBFS density forecast $p_{C,\tau}(y^*|y)$.

The results for selected years, which illustrate the logic behind the PBFS, are presented in Figure 1. Several points warrant a comment. First, forecast densities obtained from the PBFS can be nonstandard: skewness, bimodality or other irregular features may arise. Second, the PBFS output does not resemble the results of a simple forecast averaging with constant weights as the position of the combined density peak relative to the statistical and expert forecast pdfs differ across the forecasting rounds. Finally, pdf from the PBFS can have higher or lower dispersion than its individual components. In particular, if expert judgment and forecast from the statistical model do not diverge too much and are informative, i.e. they are not too similar to the unconditional forecast pdf, the PBFS yields a relatively concentrated density. On the other hand, if the expert and statistician substantially disagree about the central tendency, the forecast uncertainty implied by the PBFS is high. We consider these features of our method to combine forecasts as very appealing.

5 Out-of-sample forecasting performance

We believe that our discussion has so far demonstrated that the PBFS is a promising method that can be used to combine expert and model-based forecasts. However, a natural question arises: how does the PBFS perform relative to some benchmarks? We address this question by evaluating the out-of-sample point and interval forecasts for US inflation from the AR(1) model, the SPF and the PBFS.

For each year τ from the period 1992-2009 we generate density forecasts of annual inflation from the competing models in line with the algorithm described in section 4. The choice of the evaluation sample is motivated by the change in the number of bins and their ranges that took place in the SPF questionnaire in 1992 and made the earlier observations incomparable to the most recent ones. As a result, the evaluation sample consists of 18 observations. Forecasts for year τ are compared to actuals a_τ from the "first release" vintage, namely the advance release of the US national accounts from the first quarter of year $\tau + 1$. As discussed in Croushore (2006), experts from the SPF aim to accurately predict the first release of inflation figures by the Bureau of Economic Analysis rather than the final revised data. Consequently, it seems natural that one of the inputs to the PBFS, and hence the PBFS itself, should be evaluated with the "first release" actuals.

We start our forecasting contest by calculating point forecasts f^j_τ, which are given by the median of the density forecasts, and the corresponding forecast errors $e^j_\tau = a_\tau - f^j_\tau$, where $j = C, E, S$ for the PBFS, SPF and AR model, respectively. Standard out of sample performance statistics, i.e. the mean forecast error (MFE) and root mean squared forecast error (RMSFE), show that the SPF performs best and the AR model is the worst (see Table 1). The calculated MFEs indicate that all forecasts overestimate the level of inflation, which might be due to the downward trend in inflation in the evaluation sample. As regards the RMSFEs, the SPF is by almost 30% more accurate than the PBFS and by more than 40% better than the AR model. We analyze whether the out-of-sample performance differences suggested by the RMSFE statistics are significant with the Harvey et al. (1997) test (HLN), which corrects the Diebold and Mariano (1995) test for the sample size. The results show that the above differences are significant at the 1% level. Moreover, at the 10% level, the PBFS outperforms

5

the AR model in forecasting accuracy. It should be noted that these results should be interpreted with caution as the HLN test does not take into account the real-time nature of the data applied in the analysis (see Clark and McCracken, 2009 for a detailed exposure of the problem).

Overall, the comparison of point forecasts confirms the well-known finding from the literature, stating that point forecasts from the SPF are very difficult to beat. It should be emphasized, however, that the inferior performance of the PBFS point forecasts relative to the SPF can be at least partly explained by the fact that the AR model it incorporates is definitely too simple to account for the declining trend in inflation observed in the evaluation sample.

In the second step we analyze the performance of the competing models in terms of interval forecasts. An interval forecast can be considered well calibrated if the frequency of realized observations falling into the forecast interval equals the given coverage probability of this interval α. To test for correct interval coverage we follow Christoffersen (1998) by calculating the sequence of hit indicators:

$$I_\tau^j(\alpha) = \begin{cases} 1 \text{ if } a_\tau \in [F_{j,\tau}^{-1}(\frac{1-\alpha}{2}), F_{j,\tau}^{-1}(\frac{1+\alpha}{2})] \\ 0 \text{ if } a_\tau \notin [F_{j,\tau}^{-1}(\frac{1-\alpha}{2}), F_{j,\tau}^{-1}(\frac{1+\alpha}{2})] \end{cases} \tag{10}$$

where $F_{j,\tau}$ is the cumulative density function of forecast from method j, with $j = C, E, S$. For a correctly specified forecast, the sample estimates of interval coverage $\hat{\alpha}^j = \frac{1}{18}\sum_{\tau=1992}^{2009} I_\tau^j(\alpha)$ should be equal to their true value α.

The upper-left panel of Figure 2 shows the empirical coverage probabilities for the competing forecasting methods, plotted against their theoretical counterparts. An ideal forecast should thus have an empirical coverage rate which is located on the 45 degree line. According to our results, both the AR model and the SPF tend to generate too wide interval forecasts, which confirms the findings of Diebold et al. (1999) that the SPF tends to overestimate uncertainty.[3] The curve for the PBFS curve is closest to the 45 degree line, which indicates that the PBFS produces better interval forecasts than its two components.

[3] This is despite the fact that individual forecasters from the survey usually underestimate uncertainty of their inflation forecasts (Giordani and Soderlind, 2003).

We conduct a more formal analysis of the interval forecast accuracy with three tests proposed by Christoffersen. The first one tests the null $H_0 : E\left\{I_\tau^j(\alpha)\right\} = \alpha$, which states that unconditional coverage is correct. If the null is true, the relevant likelihood ratio test statistic LR_{uc} is $\chi^2(1)$ distributed. The results presented in the upper-right panel of Figure 2 show that $\hat{\alpha}$ from the AR model and the SPF in most cases are significantly above α. The PBFS is the only method that is able to generate forecasts with a correct unconditional coverage. As emphasized by Christoffersen, tests for the correct unconditional coverage are insufficient when dynamics are present in the data generating process of $I_\tau^j(\alpha)$. This problem is addressed by the second test with the null $H_0 : E\left\{I_\tau^j(\alpha)\right\} = E\left\{I_\tau^\alpha | I_{\tau-1}^\alpha\right\}$, which states that $I_\tau^j(\alpha)$ is independently distributed. Under the null, the relevant likelihood ratio test statistic $LR_{ind} \sim \chi^2(1)$. According to the results presented in the bottom-left panel of Figure 2, autocorrelation should not significantly influence the unconditional coverage test in our application. This is confirmed by the results of the joint likelihood ratio test of unconditional coverage and independence, where the relevant statistic under the null $LR_{cc} \sim \chi^2(2)$ (see the bottom-right panel of Figure 2).

6 Concluding remarks

The aim of this article was to design a formal method of combining expert and model-based density forecasts when the sample of past forecasts is unavailable, without resorting to ad hoc weights. This kind of framework naturally addresses the problem faced by decision makers, who often need to incorporate their judgment about the future into model-based density forecasts. The PBFS proposed in this article works directly with expert's density forecast and endogenously delivers implicit weights for forecast combination. Its output can be interpreted as a result of merging information from three sources: an expert, a statistical model and the economic theory.

In the second part of the article we have shown how to implement the PBFS algorithm to combine US inflation forecasts from a simple autoregressive model and the SPF. The results suggest that our framework can be successfully used in empirical applications. Given the relative simplicity of the autoregressive model, a natural extension of our analysis would be to apply the PBFS to combining expert forecasts with more sophisticated tools, such as dynamic stochastic general equilibrium models, which are currently used in many policy-making institutions. While such an application would certainly be much more demanding in terms of computational time, there is nothing in the PBFS preventing its use in such cases. The other possible extension would be to allow for multiple experts or multiple models. We leave these issues for future research.

References

Bernardo, J. M. and Smith, A. F. (1994). *Bayesian Theory.* John Wiley & Sons, Chichester.

Carnap, R. (1950). *Logical Foundations of Probability.* The University of Chicago Press, Chicago.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–62.

Clark, T. E. and McCracken, M. W. (2006). Forecasting of small macroeconomic VARs in the presence of instabilities. Research Working Paper RWP 06-09, Federal Reserve Bank of Kansas City.

Clark, T. E. and McCracken, M. W. (2009). Tests of equal predictive ability with real-time data. *Journal of Business & Economic Statistics*, 27(4):441–454.

Croushore, D. (1993). Introducing: the survey of professional forecasters. *Business Review*, pages 3–15.

Croushore, D. (2006). An evaluation of inflation forecasts from surveys using real-time data. Working Papers 06-19, Federal Reserve Bank of Philadelphia.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130.

de Finetti, B. (1972). *Probability, Induction and Statistics: The Art of Guessing*. John Wiley & Sons, New York.

de Finetti, B. (1975). *Theory of Probability*, volume 2. John Wiley & Sons, New York.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999). Evaluating density forecasts of inflation: The survey of professional forecasters. In Engle, R. and White, H., editors, *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, pages 76–90. Oxford University Press, Oxford.

Elliott, G. and Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics*, 122(1):47–79.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference.* Oliver & Boyd, Edinburgh.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100:680–701.

Geisser, S. (1993). *Predictive Inference: An Introduction.* Chapman & Hall, New York.

Giordani, P. and Soderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, 47(6):1037–1059.

Good, I. J. (1950). *Probability and the Weighing of Evidence.* Charles Griffin & Co., London.

Guidolin, M. and Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, 150(2):297–311.

Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1):1–13.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hill, B. M. (1985). Some subjective bayesian considerations in the selection of models. *Econometric Reviews*, 4(2):191–246.

Hill, B. M. (1992). Bayesian nonparametric prediction and statistical inference. In Goel, P. and Iyengar, N., editors, *Bayesian Analysis in Statistics and Econometrics*, pages 43–76. Springer-Verlag, Berlin.

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press, Oxford.

Jore, A. S., Mitchell, J., and Vahey, S. P. (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25(4):621–634.

Kadane, J. B. (1980). Predictive and structural methods for eliciting prior distributions. In Zellner, A., editor, *Bayesian Analysis in Econometrics and Statistics*, chapter 8, pages 89–93. North-Holland, New York.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S., and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, 75(372):845–854.

Keynes, J. M. (1921). *A Treatise on Probability.* Macmillian and Co., New York.

Klamer, A. (1984). *Conversations with Economists.* Rowman & Littlefield Publishers, Inc., Lanham.

Lane, D. A. (1985). Comment. *Econometric Reviews*, 4(2):253–258.

Lawrence, M., Goodwin, P., O'Connor, M., and Onkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3):493–518.

Lindley, D. V. (2006). *Understanding Uncertainty.* John Wiley & Sons, New Jersey.

Min, C.-K. and Zellner, A. (1993). Bayesian and non-bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, 56(1-2):89–118.

Morris, P. A. (1974). Decision analysis expert use. *Management Science*, 20:1233–1241.

Pagan, A. R. (2003). Report and modelling and forecasting at the bank of england. *Bank of England Quarterly Bulletin*, Spring.

Poirier, D. J. (1995). *Intemediate Statistics and Econometrics: A Comparative Approach*. The MIT Press, Cambridge, Massachusetts.

Rubaszek, M. and Skrzypczynski, P. (2008). On the forecasting performance of a small-scale DSGE model. *International Journal of Forecasting*, 24(3):498–512.

Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123–1175.

Stigler, S. M. (1982). Thomas Bayes's bayesian inference. *Journal of the Royal Statistical Society. Series A*, 145:250–258.

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 4, pages 135–196. Elsevier, Amsterdam.

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, 24(4):630–650.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, 67(s1):983–994.

Tables and figures

Table 1: Point forecast evaluation

Standard statistics			
	PBFS (C)	SPF (E)	AR (S)
MFE	-0.41	-0.13	-0.48
RMSFE	0.59	0.41	0.72

Harvey et al. test		
H1	HLN stat.	probability
SPF dominates PBFS	2.84	0.01
SPF dominates AR	2.56	0.01
PBFS dominates AR	1.46	0.08

Figure 1: Forecast pdfs from PBFS and its components for US inflation

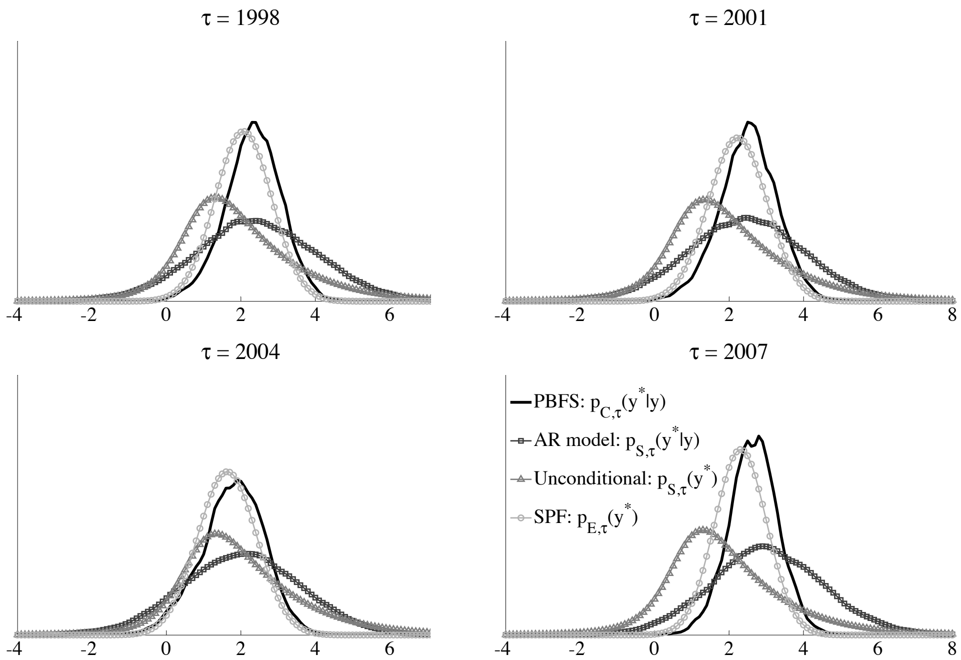

Figure 2: Interval forecast evaluation