

TI 2018-069/III
Tinbergen Institute Discussion Paper



The Evolution of Forecast Density Combinations in Economics

Knut Are Aastveit¹

James Mitchell²

Francesco Ravazzolo³

Herman van Dijk^{1,4}

¹ Norges Bank

² Warwick Business School

³ Free University of Bozen/Bolzano

⁴ Erasmus University

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

The Evolution of Forecast Density Combinations in Economics*

Knut Are Aastveit[†] James Mitchell[‡] Francesco Ravazzolo[§] Herman K. van Dijk[¶]

August 27, 2018

Abstract

Increasingly, professional forecasters and academic researchers present model-based and subjective or judgment-based forecasts in economics which are accompanied by some measure of uncertainty. In its most complete form this measure is a probability density function for future values of the variables of interest. At the same time combinations of forecast densities are being used in order to integrate information coming from several sources like experts, models and large micro-data sets. Given this increased relevance of forecast density combinations, the genesis and evolution of this approach, both inside and outside economics, is explored. A fundamental density combination equation is specified which shows that various frequentist as well as Bayesian approaches give different specific contents to this density. In its most simplistic case, it is a restricted finite mixture, giving fixed equal weights to the various individual densities. The specification of the fundamental density combination is made more flexible in recent literature. It has evolved from using simple average weights to optimized weights and then to ‘richer’ procedures that allow for time-variation, learning features and model incompleteness. The recent history and evolution of forecast density combination methods, together with their potential and benefits, are illustrated in a policy making environment of central banks.

JEL codes: C1, C15, C53, E37.

Keywords: Forecasting; Model Uncertainty; Density Combinations

*This working paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. This paper was prepared for inclusion in the Oxford Research Encyclopedia of Economics and Finance. The authors thank two referees and seminar participants at Norges Bank for helpful comments.

[†]Norges Bank and Centre for Applied Macro and Petroleum Economics, BI Norwegian Business School; Knut-Are.Aastveit@norges-bank.no

[‡]Warwick Business School, University of Warwick; James.Mitchell@wbs.ac.uk

[§]Free University of Bozen/Bolzano and Centre for Applied Macro and Petroleum Economics, BI Norwegian Business School; francesco.ravazzolo@unibz.it

[¶]Erasmus University Rotterdam, Tinbergen Institute, Norges Bank; hkvandijk@ese.eur.nl

1 Introduction

Given the increasing awareness amongst policy makers, professional forecasters and academic researchers of the uncertainties associated with any point forecast, it is now more and more common for model-based and subjective or judgment-based forecasts in economics to be accompanied by some measure of uncertainty. In its most complete form this measure is a probability density function for future values of the variable or variables of interest. Specific probability event forecasts of interest and/or confidence or credibility intervals can always be extracted from the underlying forecast density.

A leading example of a forecast density being produced and used in practice is the Bank of England's "fan chart" for GDP growth and inflation; these have been published each quarter since 1996; from 2013 the Bank of England added a fan chart for unemployment. As Figure 1 illustrates - for GDP growth - the fan charts currently show three best critical or highest posterior density regions, corresponding to thirty, sixty and ninety per cent of the forecast density. These forecast densities are not assumed symmetric so that the balances of risks may be on the upside or downside; specifically, the Bank allow the (conditional) mean forecast to differ from the modal forecast by assuming that their forecast density is a two-piece normal distribution; for a detailed discussion and evaluation of the risk forecasts produced by central banks, including the Bank of England, see Knüppel and Schulte frankenfeld (2012).

As the Bank of England's Monetary Policy Committee (MPC) explains in a note attached to each published fan chart: "If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that the mature estimate of GDP growth would lie within the darkest central band on only 30 of those occasions. The fan chart is constructed so that outturns are also expected to lie within each pair of the lighter green areas on 30 occasions. In any particular quarter of the forecast period, GDP growth is therefore expected to lie somewhere within the fan on 90

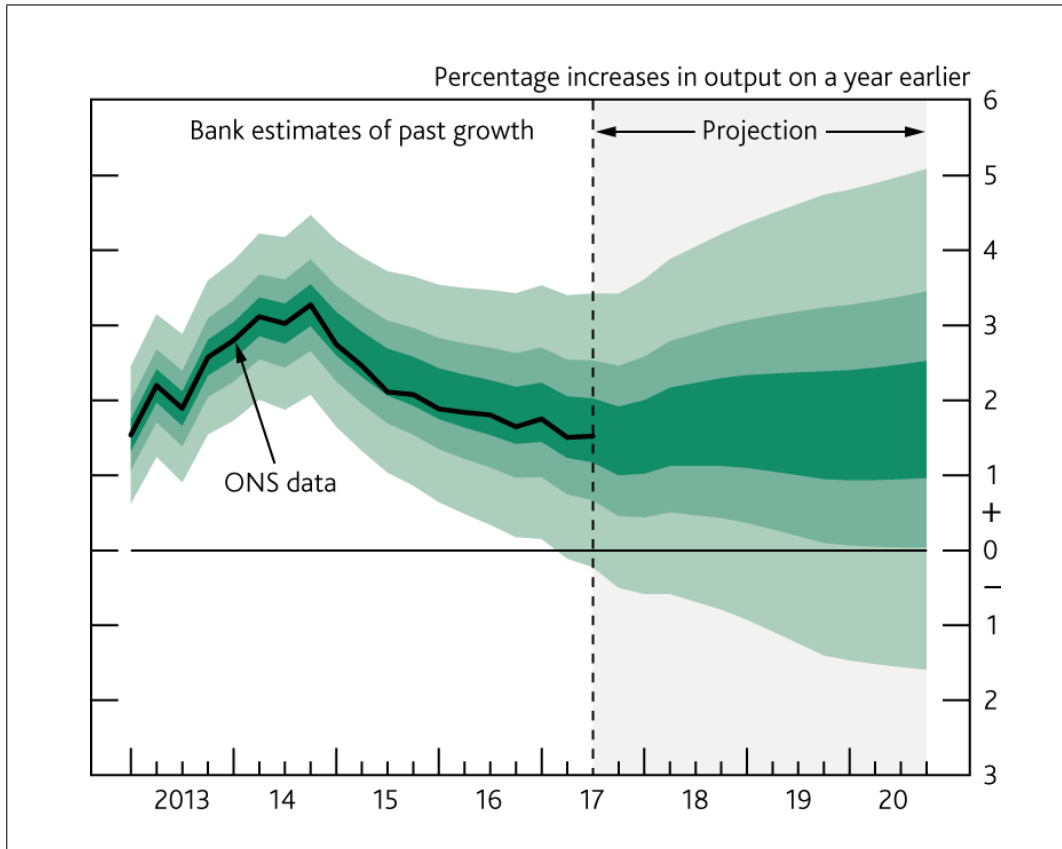


Figure 1: GDP fan chart from the Bank of England's November 2017 Inflation Report

out of 100 occasions. And on the remaining 10 out of 100 occasions GDP growth can fall anywhere outside the green area of the fan chart.”. This serves, as we discuss below, as the basis for how forecast densities can be evaluated, *ex post*, once realizations of the variable of interest have been observed. When the (economic) loss or scoring function of the user of the forecast is known, the forecast density can additionally be evaluated or scored on its basis. As Gneiting (2011) stresses, the “optimal” point forecast extracted from, for example, the MPC’s fan chart depends on the form of the user’s loss function. As is increasingly recognized, only when the user’s loss function takes a specific form - namely a Bregman function - is the optimal point forecast, in fact, the still ubiquitous (conditional) mean forecast. Measures of risk can also be extracted from forecast densities; e.g. see Lutz and Manganelli (2007) for an application quantifying the risk of deflation and Casarin et al. (2018) for applications to economic growth probabilities and Value-at-Risk

measures of stock prices.

In practice whether the forecast density is produced by a single expert (or indeed group of experts) or a single econometric model there is likely also uncertainty about whether this single forecast is “correct”. While the expert consulted or model used to produce the forecast may be the *best* forecast (from the set consulted) it may not be, indeed most likely is not, the *truth*. In many applications, certainly in economics, there is effectively “model uncertainty” with multiple density forecasts produced from different models and/or experts often available; with experience (e.g. see Rossi (2013)) often suggesting that the relative performance of different models varies over time (due to structural instabilities in the unknown underlying data-generating-process). This basic observation motivates consideration of various models¹ and the use of forecast combination methods to “integrate out” uncertainty about the best model. The goal is to produce, via combination, more accurate - and/or more useful - forecast densities, or at least density forecasts more robust to changes in the economic environment.

We structure this focused survey on the evolution of forecast density combinations as follows. First, in section 2, we review its genesis by emphasizing how the methods currently used involve the extension of well-established methods for the combination of point forecasts (that have themselves been well-reviewed previously (see Timmermann (2006))). They can also be seen to draw on and/or relate to a broad literature in management science and statistics on how to combine or reconcile the probabilistic judgments or opinions of different “experts”. Next, in section 3, we present *the fundamental combination equation* and discuss how various frequentist as well as Bayesian approaches can be seen as giving it different specific content. In its most simplistic case, the combination density is a restricted finite mixture, giving a fixed equal weight to the various individual densities. We go onto discuss how the recent literature on forecast combinations has evolved to consider making the specification of the combination density more flexible. This has involved

¹Or various “experts”, to anticipate a link we make below to a literature outside economics and finance.

going from representing it as a simple average to use of optimized weights; and then onto ‘richer’ procedures, such as allowing for time-variation in the combination weights, learning features and accommodating model incompleteness. In so-doing we both extend and update previous reviews and textbook treatments of forecast density combinations in economics by Hall and Mitchell (2009) and Elliott and Timmermann (2016); and provide an explanation and context for recent developments. Then, in section 4, we demonstrate the ‘fruits’ of this evolution process by reviewing selected applications of forecast density combination methods in a policy making environment of central banks.

Remark. In this survey we take the forecast densities to be combined essentially as given, and do not discuss in detail important issues around how the individual or component forecast densities themselves should be produced, whether via an econometric model and/or via subjective (expert) judgement. But we do stress that in economics, especially post Great Moderation and now post Global Financial Crisis, there is an increasing awareness of the utility of forecasting models that capture changes not just in the conditional mean but also in the conditional variance. As a result, models with time-varying parameters and stochastic volatility are popular, with time-varying parameter (Bayesian) Vector Autoregressive models with stochastic volatility increasingly a popular workhorse forecasting model (e.g. see Clark (2011)).

We also defer consideration of interesting and largely neglected issues around how forecast densities should best be communicated.

2 Genesis of Forecast Density Combinations

Forecast density combinations have a long, albeit scattered, history given their roots across different applied statistics fields extending into management science, with many innovative applications in meteorology. Below we firstly set out how forecast density combination methods, as increasingly seen in applied econometric applications, can be interpreted as

building on earlier work on point forecast combinations and also the (implicit) use of forecast density combinations when presenting results from surveys of professional forecasters that explicitly seek to quantify forecasters' uncertainty. Looking beyond economics, and anticipating more recent Bayesian approaches to forecast density combinations below (McAlinn and West (2018), Billio et al. (2013) and Aastveit et al. (2018)), we summarize a relevant literature in management science, meteorology and mathematical statistics on combining experts' probabilistic predictions to arrive at what the management science literature often call a "consensus" distribution.

2.1 Point forecast combinations

In the financial press and in academic research (e.g. see Batchelor and Dua (1995) and Ager et al. (2009)) frequent mention and use is made to the "Blue Chip Average Forecast" and to the "Consensus Forecast". These are averages of the multiple (point) forecasts of professional forecasters of the same (macroeconomic or financial) variable. And empirical experience, as reviewed in Timmermann (2006), is that these type of combined forecasts often work well in practice in economics. As Timmermann (2006) reviews, the success of combinations follows from the fact that individual forecasts may be based on misspecified models, poor estimation or non-stationarities. Their use also offers one way to acknowledge forecast uncertainty absent direct measures of forecast uncertainty. But this is a partial acknowledgement, since forecast disagreement is only one component of forecast uncertainty; e.g. see Wallis (2005) for a discussion, appropriately for the present paper, in the context of forecast density combinations of the sort used by the SPF (considered in section 2.2.1 below).

The foundational paper used to rationalize implementation of point forecast combination is Bates and Granger (1969), where the authors motivate combination via a portfolio diversification argument. As the information sets behind the individual forecasts are often

unknown (in the language subsequently used below, they are unknown to the “decision maker” undertaking the combination) it is often not possible to combine them. But the forecasts of y_t (for time $t = 1, \dots, T$) from the different forecasters or experts, denoted \tilde{y}_{it} (for forecast or “expert” $i = 1, \dots, N$), can be combined directly. Bates and Granger (1969) make use of a weighted linear combination of the form:

$$\sum_{i=1}^N w_{it} \tilde{y}_{it} \tag{1}$$

where the w_{it} are the combination weights. These weights can in principle vary both across experts and time.

This weighted average is intended to be a good approximation to y_t . As Bates and Granger (1969) explain, the error variance (or Root Mean Squared Error (RMSE)) of the combined forecast, (1), is minimized by setting the combination weights w_{it} equal to the least squares estimates of w_{it} derived from a linear regression of the (subsequent) realization y_t on the N point forecasts, \tilde{y}_{it} , subject to the restriction that $\sum_{i=1}^N w_{it} = 1$; see also Granger and Ramanathan (1984) who consider extensions that both allow for biased (point) forecasts, by including an intercept in the linear regression, and for unrestricted regression coefficients. Such a combined point forecast is “optimal” - in the specific sense that it minimizes the RMSE loss function.

Many extensions and refinements of Bates and Granger (1969) have been suggested. For example, Terui and van Dijk (2002) generalize the least squares weights by representing the dynamic forecast combination as a state space model with weights that are assumed to follow a random walk process. Guidolin and Timmermann (2009) introduce Markov-switching weights, and Hoogerheide et al. (2010) propose robust time-varying weights and account for both model and parameter uncertainty in model averaging. Raftery et al. (2010) derive time-varying weights in a “dynamic model averaging” framework, following in the spirit of Terui and van Dijk (2002), and they speed up com-

putations by applying forgetting factors for the recursive Kalman filter updating. And there have been numerous applications with, for example, Clark and McCracken (2010) finding that linear combinations of point forecasts, as produced in real-time, are an effective means of mitigating the structural or what they call “uncertain instabilities” that are found to plague individual macroeconomic forecasting models; and Stock and Watson (2004) document the robust performance of point forecast combinations for numerous economic variables. Timmermann (2006) and Elliott and Timmermann (2016) provide comprehensive reviews of this large literature.

One stylized fact or puzzle that has been drawn out from this point forecast combination literature is that, in practice, it is often hard to beat an equal weighted combination, where $w_{it} = 1/N$, despite the fact that Bates and Granger (1969) weights are, in theory, “optimal” - as stressed above, optimal in terms of minimizing RMSE. Recent studies have sought to rationalize this empirical finding. Smith and Wallis (2009) consider that use of Bates and Granger (1969) weights involves parameter estimation error; and it is this estimation error that may obscure any benefits of optimized weights. While Elliott (2017) explains the puzzle with reference to the predictability of the series to be forecasted, showing that the larger the unpredictable component, the smaller the potential gains from using optimized weights that involve estimating correlations from the data.

Providing a bridge between point forecast combination and more recent applications of forecast density combinations in economics are Clements (2002) and Granger et al. (1989) who consider, respectively, the combination of multiple event and quantile forecasts. While these methods inevitably involve a loss of information compared with consideration of the ‘whole’ density, they are a precursor.

2.2 Early uses of combined forecast densities

Here we consider what we dub ‘early’ applications and uses of combined density forecasts both within and outside economics. These focus on use of the so-called “linear opinion pool” to aggregate the probabilistic forecasts, as the linear opinion pool can be seen as a generalization of the linear point forecast combinations seen in (1). We go onto explain that the forecast density combination literature has evolved, in particular in economics, to consider alternatives to linear combinations.

2.2.1 In economics: the Survey of Professional Forecasters

As Wallis (2005) stresses, the Survey of Professional Forecasters (SPF) in the US, previously the ASA-NBER survey, has, in effect, used a combined forecast density (or a finite mixture distribution) to publish its much-analyzed aggregate forecast densities for inflation and GDP growth since 1968. In the spirit of (1), the SPF takes a linear combination of N different forecast densities (each professional forecaster or “expert” produces a forecast density $p(y_t|I_i)$ via the “linear opinion pool”:

$$p(y_t|I) = \sum_{i=1}^N w_{it}p(y_t|I_i) \tag{2}$$

where w_{it} is the weight given to forecast density i ($i = 1, \dots, N$), I_i is the information set belonging to model i and I is the joint information set across all N models.² Here we assume the $i = 1, \dots, N$ individual forecast densities, $p(y_t|I_i)$, are continuous; and we abstract from the fact that, at least in many surveys like the US SPF, these forecast densities $p(y_t|I_i)$ are usually known in discrete form - as respondents to the SPF actually provide histogram forecasts rather than continuous forecast densities. As we elaborate in section 3 below, and to provide a link back to (1), we might interpret \tilde{y}_{it} , in a simulation

²We do not distinguish or discuss issues that arise from the density forecasts being produced at different forecast horizons; i.e. we do not specify what time period I_i refers to. Instead we simply note that, for example, a one-period ahead density forecast would be denoted $p(y_t|I_{t-1})$.

context for example, as comprising draws from $p(y_t|I_i)$.

As the combined density is a linear combination of the N individual densities, the variance of the combined forecast density will generally be higher than that of the individual models. This is because the variance of $p(y_t|I)$ is a weighted sum of a measure of model uncertainty and dispersion of (or disagreement about) the point forecast (e.g. see Wallis (2005)).

When publishing the aggregated forecast densities from the US SPF it is assumed that $w_{it} = w_i = \frac{1}{N}$. Despite the long history of the SPF, little attention (until later evolutions in economics that we consider below) has historically been paid to how the weights on the competing forecast densities in the finite mixture should be determined. As experience of combining point forecasts has taught us, irrespective of its performance in practice, use of equal weights is only one of many options. For example, as discussed in section 2.1, one popular alternative to equal weights in the point forecast literature, the so-called regression approach, is to tune the weights to reflect the historical performance of the competing forecasts - and to weight more highly those forecasts that historically have been “good” forecasts. Such an approach is popular in many applications in economics where the sample is (often, but not always) long enough to allow for a “training sample” so that the weights can, in principle, be tuned to reflect the past performance of the N “experts”.

One rationale for forecast density combinations is seen by noting that (2), being a linear combination of the N densities, can generate a combined density with properties quite distinct from those of the N component densities. For example, if all the component densities are normal then the combined density will be mixture normal. Mixture normal distributions can have heavier tails than normal distributions, and can potentially accommodate skewness and kurtosis. If the true (population) density is non-normal, one can begin to appreciate why combining individual forecast densities, that are normal, may mitigate misspecification of the individual densities. Equally, if the true distribution is

normal combining using (2) will, usually, not be a good approximation strategy.

2.2.2 A digression outside economics

We take a brief digression in order to connect with important developments in related fields like management science, meteorology and mathematical statistics on combining predictive densities.

Management science and risk analysis The combination of predictive densities has received considerable attention within management science and risk analysis dating back, at least, to Winkler (1968); for reviews see Genest and Zidek (1986) and Clemen and Winkler (1999). The often-found forecasting gains from aggregation of multiple forecasts is often referred to as the “wisdom of the crowd”.

Clemen and Winkler (1999) distinguish behavioral and mathematical approaches to forecast density combination. The behavioral approach seeks to combine experts’ opinions by letting the experts interact in some manner to reach some collective opinion. This approach is not considered further here, given that the attention in economic forecasting is often focused on combining model based forecasts. By contrast, mathematical statistical approaches, often, combine the information across experts by using some formal rule. Early work in this vein focused on combination rules that satisfied certain properties or axioms. Common axiomatic approaches are the “linear opinion pool” and the “logarithmic opinion pool”. Second to the linear opinion pool, the logarithmic opinion pool is the most widely used density forecast combination method in economics; e.g. see Kascha and Ravazzolo (2010) and Wallis (2011).

A Bayesian approach to forecast density combination also has deep roots, whereby experts’ densities are combined by a “decision maker” who views them as data. Following Winkler (1968), Morris (1974) and Morris (1977) and anticipating our discussion of McAlinn and West (2018) in section 3.4.2, Bayes’ Theorem is used to update the decision

maker's prior distribution in the light of these data from the experts that takes the form of the joint density, or likelihood, derived from their N densities. The difficulty faced by the decision maker is deciding upon the form of the likelihood function; the likelihood must capture the bias and precision of the experts' densities as well as their dependence. This initially hindered use of this approach in economics. Recent developments, which have overcome several of these limitations, are discussed later in this survey; see Section 3.4 and our discussion of flexible Bayesian combination structures.

To proceed - and render the Bayesian approach tractable - assumptions need to be made; e.g. Winkler (1968) considers the prior density to be a member of a natural-conjugate family of distributions, so that the posterior - or combined - density is of the same form. Interestingly, Winkler (1968) compares such a combined forecast density with what he calls the Weighted-Average approach which is (2) and emphasizes the differences between the two forecast density combination strategies. Winkler (1981) derives tractable expressions for the combined or posterior forecast density based on the multivariate normal distribution, that importantly explicitly accommodates dependence between the different experts' forecast densities by measuring the dependence (defined by correlation) between the experts' forecasting errors. The posterior (or combined) mean forecast from Winkler (1981) is equivalent to the minimum variance solution of Bates and Granger (1969) concerned with the combination of point forecasts accommodating their (linear) dependence. The copula approach proposed by Jouini and Clemen (1996), and considered further in Mitchell (2013) in a macroeconomic application to fan charts, takes up further the issue of how to accommodate expert dependence, defined more generally than correlation, when combining their forecast densities. The Bayesian approach has recently become popular in empirical economics. We take this further in section 3.

Meteorology and mathematical statistics The meteorology literature has also been at the vanguard of combinations of probability forecasts, with applications from Sanders

(1963) through to a range of more recent papers including Gneiting et al. (2005), Sloughter et al. (2010) and Ranjan and Gneiting (2010). A point of departure in many of these applications reflects the fact that many meteorological forecasting systems deliver (multiple, depending on initial conditions) point forecasts rather than multiple forecast densities *per se*. So, instead of constructing the combined forecast density as some aggregation, such as the linear combination in (2), of N different forecast densities, assumed known, a combined or “ensemble” density is constructed from a collection of point forecasts (e.g. see Raftery et al. (2005) and Gneiting et al. (2005)). While such methods have, to-date, received less attention in economics than combinations of N *known* forecast densities, Krüger (2017) provides a nice application illustrating the utility of these methods when combining the GDP growth and inflation point forecasts of “experts” from the European Central Bank’s SPF.

In a series of papers in what might be broadly called mathematical statistics (e.g. see Ranjan and Gneiting (2010)), but also echoing a recent work stream in management science (e.g. see Hora (2004)), the properties of linearly combined forecast densities have been analysed; and alternatives to the linear opinion pool have been proposed. Some of these are discussed further below in section 3.2. From an economics’ perspective, important takeaways from the theoretical contributions of this work are:

1. A linear combination, (2), of N individually well-calibrated forecast densities must be uncalibrated and lack sharpness, as the linear opinion pool blows up the variance of the combination (see Hora (2004) and Ranjan and Gneiting (2010)). Or, in other words, the performance of the combined forecast density depends on how well calibrated are the individual $i = 1, \dots, N$ forecast densities. When over-confident, in the sense that the experts deliver predictive densities that are too narrow (so that, for example, fewer than 90% of realizations fall within a 90% forecast interval), we can imagine that the linear opinion pool might be helpful. For further discussion,

theoretical results and motivation for alternatives to the linear opinion pool such as the “trimmed linear pool”, that seek to trim away from the combination forecasts with low or high means, or cumulative distribution function values, in order to decrease the variance, see Grushka-Cockayne et al. (2017).³ Quantile aggregation has also received some attention, with a macroeconomic application illustrating its utility provided by Buseti (2017); quantile aggregation involves taking a (perhaps weighted) linear combination of the $i = 1, \dots, N$ quantile functions rather than their inverse as in the linear opinion pool. For further analysis within management science/decision analysis see, for example, Lichtendahl Jr. et al. (2013) and Hora et al. (2013). Interestingly, unlike (2), quantile aggregation means that the combined density $p(y_t|I)$ belongs to the same family as the individual densities $p(y_t|I_i)$ assuming these are all from the same location-scale (e.g. Gaussian) family.

2. Nonlinear pools, such as the beta-transformed linear pool of Ranjan and Gneiting (2010), and the generalized and spread-adjusted pools of Gneiting and Ranjan (2013), can be preferable to linear pools in terms of delivering better calibrated forecast density combinations. Anticipating the discussion in section 3, these approaches can be seen as specific instances of the fundamental density combination approach. For a full discussion in a Bayesian context, and the development of a flexible nonparametric forecast density combination approach using beta mixtures that allows for model “incompleteness” (discussed further in section 3) see Bassetti et al. (2018), who also provide a finance application forecasting S&P returns.
3. As an alternative to letting the form of the combination depend on its components, as it does in the linear opinion pool, one can combine by imposing a specific form for the combined density forecast but combine component information (say the predictive means) rather than the predictive densities per se. This is what variants

³As Grushka-Cockayne et al. (2017) discuss, psychologists have shown that (human) experts have a general tendency to be overconfident in the forecasts they produce.

of the so-called EMOS approach used more widely in meteorology do; e.g. see Gneiting et al. (2005) who assume the combined density is normal.

3 Recent Forecast Density Combinations in Economics

3.1 The Fundamental Density Combination Equation

Rather than present a chronological analysis of different forecast density combination methods, we organize our discussion around two issues that have emerged as the recent literature on this topic in economics has evolved.

1. *Weighting.* How are or should different forecast densities be weighted in the chosen combination scheme; and if/how should these weights update over time, e.g. via learning, and reflect cross-forecast dependencies?
2. *Incompleteness.* How should, and how many, density forecasts should be combined to accommodate likely model incompleteness and model misspecification?

These two issues have been dealt with in various ways in the literature. We focus our discussion of these issues by introducing the following central concept - namely, the fundamental density combination equation. For a basic development we refer to Billio et al. (2013) and for a related formal Bayesian foundational motivation see McAlinn and West (2018) that we also summarize in section 3.4.2 below.

Fundamental density combination equation We write the fundamental forecast density combination equation for $p(y_t|I)$ as

$$p(y_t|I) = \int p(y_t, \tilde{y}_t|I)d\tilde{y}_t = \int p(y_t|\tilde{y}_t)p(\tilde{y}_t|I)d\tilde{y}_t, \quad (3)$$

where $\tilde{y}'_t = (\tilde{y}_{1,t}, \dots, \tilde{y}_{N,t})$ are the predicted values from the $i = 1, \dots, N$ models (or experts). We might interpret the \tilde{y}_{it} as either latent variables that relate to the variable of interest, y_t , via the conditional density $p(y_t|\tilde{y}_t)$, or in a simulation context as draws from the individual density forecasts, $p(y_t|I_i)$; see also section 3.4.2 below.

Equation (3) is, of course, an elementary one in distribution theory. $p(y_t|I)$ is a convolution of two densities. More specifically, it is a *mixture density* where $p(\tilde{y}_t|I)$ is the mixing density, which is the joint density of the $i = 1, \dots, N$ predictive densities $p(\tilde{y}_{it}|I_i)$ associated with the individual (marginal) models. Thus, we have an N -dimensional integral on the right-hand-side of equation (3). We interpret the density $p(y_t|\tilde{y}_t)$ as the *fundamental combination density*.

Various frequentist as well as Bayesian approaches to forecast density combination can be seen as variants or special cases of equation (3). Specifically, the different forecast density combination approaches can be seen to give different content to the fundamental combination density, i.e., they involve different specifications of $p(y_t|\tilde{y}_t)$. In its most simplistic case, the combination density is a restricted finite mixture, giving fixed equal weights to the various individual densities. That is, $p(y_t|\tilde{y}_t)$ can be re-interpreted as defining the model combination weights w_{it} in (2). Recent evolutions of forecast density combination methods in the literature have sought to make the specification of the $p(y_t|\tilde{y}_t)$ density ‘richer’ and more flexible. That is, in effect they have gone from using $p(y_t|\tilde{y}_t)$ to define simple average weights to using it to define optimized weights. Most recently, ‘richer’ weighting procedures, such as allowing for random latent combination weights that possess time-varying and learning features, and model incompleteness have been accommodated by considering more flexible representations for $p(y_t|\tilde{y}_t)$. A focused survey on the literature on optimized weights, which is usually based on frequentist methods, is given in the next subsection. Methods that allow for flexible updating (learning) of the combination weights are increasingly based on Bayesian methods. This topic is covered later in Section 3.4.

3.2 Frequentist based optimized combination weights

How we measure the accuracy of forecasts is central to how we choose to combine them “optimally”. Point forecasts are traditionally evaluated on the basis of their RMSE relative to the subsequent realizations of the variable; and, as discussed in section 2.1, this leads to optimal weights being defined via a least squares regression of the realizations of the variable on the competing point forecasts.

To combine density forecasts optimally similarly requires the loss function to be chosen. The starting point for Hall and Mitchell (2007) in a macroeconomic application, that involved combining Bank of England density forecasts with competing density forecasts, was the desire to obtain the most “accurate” density forecast in a statistical sense. This can be contrasted with economic approaches to evaluation, that evaluate density forecasts in terms of their implied economic value or utility when used to inform decisions made; e.g. see Granger and Pesaran (2000) and Clements (2004). The *Remark* below provides some more context on the evaluation of density forecasts.

Hall and Mitchell (2007) define the optimal weights as that set of weights, $w_{it} = w_i$, in (2) that minimize the Kullback-Leibler information criterion [*KLIC*] distance between the combined density forecast and the true but unknown density of the variable to be forecast. Practically, and conveniently, this minimization can be achieved using the logarithmic scoring rule.

Specifically, the *KLIC* distance between the true density $f(y_t)$ and the combined density forecast $p(y_t|I)$ ($t = 1, \dots, T$) is defined as:

$$KLIC_t = \int f(y_t) \ln \left\{ \frac{f(y_t)}{p(y_t|I)} \right\} dy_t \text{ or} \quad (4)$$

$$KLIC_t = \mathbf{E} [\ln f(y_t) - \ln p(y_t|I)]. \quad (5)$$

The smaller this distance the closer the density forecast to the true density. $KLIC_t = 0$

if and only if $f(y_t) = p(y_t|I)$.

Under some regularity conditions $E[\ln f(y_t) - \ln p(y_t|I)]$ can be consistently estimated by \overline{KLIC} , the average of the sample information on $f(y_t)$ and $p(y_t|I)$ ($t = 1, \dots, T$):

$$\overline{KLIC} = \frac{1}{T} \sum_{t=1}^T [\ln f(y_t) - \ln p(y_t|I)]. \quad (6)$$

Definition 1 *The optimal combined density forecast is $p^*(y_t|I) = \sum_{i=1}^N w_i^* p(y_t|I_i)$, where the optimal weight vector \mathbf{w}^* , $\mathbf{w}^* = (w_1^*, \dots, w_N^*)$, minimizes the KLIC distance between the combined and true density, (6). This minimization is achieved as follows:*

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \frac{1}{T} \sum_{t=1}^T \ln p(y_t|I), \quad (7)$$

where $\frac{1}{T} \sum_{t=1}^T \ln p(y_t|I)$ is the average logarithmic score of the combined density forecast over the sample $t = 1, \dots, T$.

While choosing the combination weights to maximize the logarithmic score has the attraction that it can be interpreted as minimizing \overline{KLIC} ; only when $\overline{KLIC} = 0$ are the combined density forecasts $p(y_t|I)$ “well-calibrated” in an absolute sense.

Pauwels and Vasnev (2016) provide further analysis of (7) and document the properties of this optimization problem; Conflitti et al. (2015) provide an iterative algorithm for (7) that enables the optimal combination to be viable even when the number of density forecasts, N , combined is large. Opschoor et al. (2017) extend the framework to consider combining density forecasts to maximize performance in a specific region of the density so as to, for example, best measure downside risk (VaR) in equity markets, based on use of a censored scoring rule and the CRPS; and Raftery et al. (2005) consider the combination of forecast densities not to maximize the logarithmic score but to minimize the CRPS instead. And in a portfolio choice application Pettenuzzo and Ravazzolo (2016) consider density forecast combinations, albeit using the Bayesian approach of Billio et al. (2013)

considered in section 3 below, where the combination weights are tuned to depend on the history of the models' past profitability.

Remark 2 : Well-calibrated density forecasts. *Well-calibrated density forecasts are defined as density forecasts such that the sequence of probability integral transforms, z_t , associated with the density forecast $p(y_t|I)$ are such that the z_t*

$$z_t = P(y_t|I) \tag{8}$$

where $P(y_t|I)$ is the cumulative distribution function associated with $p(y_t|I)$, are independent (for 1 step ahead forecasts only) uniform $U[0,1]$ values. As emphasized by Diebold et al. (1998), only when the z_t satisfy this condition will the predictive density $p(y_t|I)$ be preferred by all users of the forecast, irrespective of their (economic) loss function. This is convenient given that it is often hard in macroeconomics to define an appropriate general (economic) loss function; although in finance, in some applications, the economic loss function can be parameterized and then directly minimized and in effect used to derive an optimal portfolio allocation; e.g. see Barberis (2000) and Pettenuzzo and Ravazzolo (2016). In practice, therefore, combined density forecasts can be evaluated in an absolute sense via application of some goodness-of-fit test to assess whether the z_t are uniform and independent (for 1 step ahead forecasts); see Corradi and Swanson (2006), Mitchell and Wallis (2011) and Knüppel (2015) (and references therein) for further discussion specifically with macroeconomic (time-series) applications in mind. In turn, relative accuracy can be tested via comparison of two or more density forecasts having scored their accuracy by assigning each a numerical score based on the density forecast and the subsequent realization of the variable. Popular choices, in macroeconomics and finance, are to test relative density forecast accuracy based on differences in logarithmic scores (e.g. see Amisano and Giacomini (2007) and Mitchell and Hall (2005)).

No one combination or aggregation method has gained universal dominance in macroe-

economics and finance. Indeed, most applications still focus on the linear opinion pool albeit as we emphasize below allowing for model incompleteness and time-varying combination weights.

In concluding this section we therefore do not provide detail on alternatives to linear pooling; but we refer back to our earlier discussion in section 2.2.2 for a summary of some of these alternatives with a view to encouraging their greater use in economics.⁴

3.3 Bayesian Model Averaging weights

Bayesian Model Averaging (BMA) offers a conceptually elegant means of dealing with model uncertainty and accordingly of producing forecast density combinations; e.g. see Roberts (1965) and Draper (1995). When using BMA the individual predictive densities, $p(y_t|I_i) = p(y_t|M_i, I)$, from model M_i are combined into a combined or posterior predictive density $p(y_t|I)$, given as

$$p(y_t|I) = \sum_{i=1}^N P(M_i|I) p(y_t|M_i, I) \quad (9)$$

where the weights are specified as the posterior probability of model i , derived by Bayes' rule, and given as

$$P(M_i|I) = \frac{P(I|M_i) P(M_i)}{\sum_{i=1}^N P(I|M_i) P(M_i)}, \quad i = 1, \dots, N \quad (10)$$

and $P(M_i)$ is the prior probability of model M_i , with $P(I|M_i)$ denoting the corresponding marginal (data) likelihood.

Comparing equation (9) with (2), (superficial, as we discuss shortly) similarities are seen, given that (9) resembles a linear opinion pool where $w_{it} = \frac{P(I|M_i)P(M_i)}{\sum_{i=1}^N P(I|M_i)P(M_i)}$.

It is important to note that these posterior probabilities, seen in (10), indicate the prob-

⁴See Galvao et al. (2018) for a comparison of these approaches.

ability that model M_i is the best model in a KLIC sense (see, e.g., Fernandez-Villaverde and Rubio-Ramirez (2004)). Other work has used approximate BMA weights, in effect, to produce a combined density forecast; e.g. see Garratt et al. (2003). In a similar vein, essentially assuming an equal prior weight on each model, Jore et al. (2010) proposed to weight each forecast density based on its recursively updated average logarithmic score. Garratt et al. (2011) and Aastveit et al. (2014) find that the recursive weighting scheme of Jore et al. (2010) performs well when combining macroeconomic forecast densities. It is also important to note that the logarithmic score of the combined forecast density is not necessarily maximized when approximate Bayesian e.g. Jore et al. (2010)-type combination weights are used.

Amisano and Geweke (2010) explain that the theoretical differences between the optimized weights in (7) and the BMA weights in (10) arise from the belief in BMA that the model space is “complete”, i.e. one of the N models under consideration is correct and equals $p(y_t|I)$. Under this belief, asymptotically the posterior probabilities converge to zero, except for one which converges to unity. So, one model within the set receives all the weight and there is therefore effectively model or forecast selection rather than combination. Importantly when the model space is “incomplete” this is not a property shared by optimal weights as defined in (7); optimal weights need not converge to zero or unity and it is possible, even as $T \rightarrow \infty$, for an incorrect model to have a non-zero weight as defined by (7).

Amisano and Geweke (2010) and Mitchell and Wallis (2011) discuss apparent similarities between the combined forecast density - represented in (2) via the linear opinion pool - and model “mixtures”. It is important to note that in mixture models the weights w_{it} are assumed to be latent (0-1) random variables which sum to one and in effect involve selecting one of the i models at time t ; in regime-switching models these indicator variables often have a Markov structure. While forecast density combination involves, effectively at step one, estimating the parameters in the component forecasting models

and then at step two combining these N predictive densities, mixture estimation involves estimating both component models and the weights simultaneously. Restricted variants of this have been considered in the aforementioned meteorology literature, with Raftery et al. (2005) an example of an approach that seeks to estimate, within a BMA framework, the combination weights w_i simultaneously with selected parameters of the component densities. And Maheu and McCurdy (2009) provide a good example of the use of Bayesian methods to estimate a mixture-of-normals model, that also as discussed above accommodate structural breaks of the sort that characterize many macroeconomic and financial time-series, to give non-Gaussian density forecasts for equity returns. This topic will be pursued more in the next section.

Although several papers have found that BMA is useful for improving predictability, it suffers from several important drawbacks.

First, and perhaps most importantly, BMA assumes that the *true model* is included in the model set. As stated above, the combination weights in equation (9) converge, under such an assumption, in the limit, when the number of observations tends to infinity, to select the true model. However, all models could be false, and as a result the model set could be misspecified. This is really the issue of model incompleteness.

Secondly, BMA assumes that the model probabilities are fixed and does not account for the *uncertainty of the weights* attached to each model. In a forecasting environment characterized with large model uncertainty and large instability in the various models' performances, the weight uncertainty can be very large. It is also well-known that BMA under vague prior information is extremely sensitive. It is shown in Bastürk et al. (2017) that use of the predictive likelihood may remedy some of these shortcomings of BMA.

3.4 Flexible Bayesian forecast combination structures

We present a summary of a forecast density combination approach with time-varying learning weights and an allowance for model incompleteness, as recently developed in a series of papers; see Billio et al. (2013), Casarin et al. (2015), Aastveit et al. (2018) and Bastürk et al. (2018). For expository purposes we simplify the exposition; for details reference is made to the aforementioned papers. In the next subsection we connect again with the literature on Bayesian predictive synthesis as discussed in McAlinn and West (2018). We end this section with two comments.

3.4.1 Forecast density combinations with time-varying learning weights and model incompleteness.

We start by simply remarking that in equations (1) and (2) the combination weights are explicitly seen, while in the fundamental density combination equation, (3), the connection between y_t and \tilde{y}_t is not given a specific form or content. Therefore, we now include in equation (3) a weight density which connects the combination density with the predictive densities from the N models. Let $p(w_t|\tilde{y}_t)$ be such a continuous weight density. Then one can generalize the fundamental model, given in equation (3), to be a convolution of three densities, given as:

$$p(y_t|I) = \int \int p(y_t, w_t, \tilde{y}_t|I) dw_t d\tilde{y}_t = \int \int p(y_t|w_t, \tilde{y}_t) p(w_t|\tilde{y}_t) p(\tilde{y}_t|I) dw_t d\tilde{y}_t, \quad (11)$$

where $p(y_t|w_t, \tilde{y}_t)$ is now specified as a combination density that explicitly incorporates the weights, $p(w_t|\tilde{y}_t)$ is the weight density and $p(\tilde{y}_t|I)$ is the joint predictive density of all N models. Note that integrals are of dimension N .

As a next step, we give content to the combination density and the weight density.

We start by specifying the combination density as a normal density:

$$p(y_t|w_t, \tilde{y}_t) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left\{ -\frac{(y_t - \tilde{y}_t'w_t)'(y_t - \tilde{y}_t'w_t)}{2\sigma_\varepsilon^2} \right\}. \quad (12)$$

This implies that there exists a model that presents the connection between the N predictions from the different sources, \tilde{y}_t with y_t as:

$$y_t = \tilde{y}_t'w_t + \varepsilon_t, \text{ :: } \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \text{ :: } t = 1, \dots, T. \quad (13)$$

Compared to equation (3), the model in equation (13) contains two fundamental generalizations.

First, the vector of weights $w_t' = (w_{1t}, \dots, w_{Nt})$ consists of (*unobserved*) *random variables* so that we can model and evaluate their uncertainty. Note that one can also evaluate the correlations between the weights of the different models.

Secondly, we have added an error term ε_t which is an indication that *model incompleteness* can be modeled and evaluated. That is, as well as Bayesian learning, (12) and (13) also allow for Bayesian diagnostic analysis of misspecification.

Next, we specify the weight density. Note that these weights constitute a convex combination and are restricted to the unit interval, so that they can be interpreted as probabilistic weights. However, combination weights in economics need not be restricted to the unit interval. That is, simple positive and/or negative feedback mechanisms imply weights that range over the entire real line. Therefore, we specify a nonlinear transformation:

$$w_{it} = g(x_{it}), \quad (14)$$

where the function $g(\cdot)$ might be given as the logistic function:

$$w_{it} = \frac{\exp(x_{it})}{\sum_{i=1}^n \exp(x_{it})}, \text{ :: } i = 1, \dots, N, \quad (15)$$

and the x_{it} are random variables that have a latent component and a component that is based on past feedback mechanisms. These may contain valuable past statistical and economic information about how the weights have evolved over time.

Thus, in modelling the dynamics for the weights, we observe that these weights come, mostly, from past predictive performance or relevant economic information where the element of learning comes into consideration. This is specified for the N -vector of x_{it} as:

$$x_t = x_{t-1} + h(z_t) + \eta_t, :: \eta_t \sim NID(0, \sigma_\eta^2), \quad (16)$$

where z_t may be included to capture (observed) economic variables believed to help explain x_t . We note that Casarin et al. (2015) also use time varying variances for the disturbances of equation (16). Del Negro et al. (2016) consider a variant of this approach whereby they estimate time-varying weights in linear prediction pools, what they call Dynamic Pools, and control the prior persistence, volatility and the mean of the weight process through hyperparameters that are estimated in real-time. They illustrate the utility of time-varying weights in an application that investigates the relative forecasting performance of DSGE models with and without financial frictions for output growth and inflation from 1992 to 2011.

The Markov assumption states that the present value of x_t is only dependent on the recent past x_{t-1} not on previous past values. More detailed specifications and other examples of (16) are considered in the aforementioned papers.

For general forms of the weight and combination densities, it is usually not known how to evaluate these densities numerically. It is, for instance, not known how to generate draws directly from the weight distribution and the combined predictive density model. In this situation one can make use of representation results, due to Billio et al. (2013), which state that the density combination model can be written as a nonlinear state space model. A summary of this result is presented for the continuous case in Figure 1. This figure shows

Table 1: The combined forecast density as a nonlinear state space model.

DECO: $p(y_t I) = \int \int p(y_t w_t, \tilde{y}_t, I)p(w_t \tilde{y}_t, I)p(\tilde{y}_t I)dw_t d\tilde{y}_t.$	
Combination density	Measurement equation
$y_t w_t, \tilde{y}_t \sim N(\tilde{y}_t'w_t, \sigma_\varepsilon^2)$	$y_t = \tilde{y}_t'w_t + \varepsilon_t, :: \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$
Weight density	Link function
$p(w_t) = p(g(x_t))$	$w_{it} = \frac{\exp(x_{it})}{\sum_{i=1}^n \exp(x_{it})}$
Markov process	Transition equation
$x_t x_{t-1} \sim N(x_{t-1}, \sigma_\eta^2)$	$x_t = x_{t-1} + \eta_t, :: \eta_t \sim NID(0, \sigma_\eta^2)$

Note: The DECO equation refers to (11), the combination density and measurement equation are given by (12) and (13). The weight density and link function are given by (14) and (15). Finally the dynamics of the weights (transition equations), assumed to follow a Markow process is given by (16).

how the combined predictive density approach is connected to filtering methods seen in the literature on nonlinear and non-Gaussian modelling and inference. This approach can be used in order to evaluate numerically the combined predictive density $p(y_t|I)$.

Casarin et al. (2018) restated the continuous case and provide a representation of the forecast density combination as a large *finite* mixture of convolutions of densities from different models. The essential step is that the combination density (12) is now replaced by a finite mixture density like equation (2) where the weights are also allowed to be dependent over time and between models. This makes use of the mixture of experts' approach from Jacobs et al. (1991), Jordan and Jacobs (1994), Jordan and Xu (1995) and Peng et al. (1996). For details on the implied algorithm we refer to Casarin et al. (2018).

3.4.2 Bayesian predictive synthesis

The Bayesian predictive synthesis (BPS) framework of McAlinn and West (2018) provides a foundational Bayesian perspective for forecast density combination, based on

earlier work in agent opinion analysis by West (1992).⁵ As in the management science literature reviewed in section 2.2.2, the general starting point is that a Bayesian decision maker \mathcal{D} receives forecast densities for the variable of interest, y_t , from a set of N *agents*, labelled \mathcal{A}_i where $i = 1, \dots, N$ (these *agents* might be interpreted, as previously, as either models, surveys or indeed experts). These forecast densities represent the individual inferences from the agents, they are viewed as data and define the information set $I = \{p(y_t|I_1), \dots, p(y_t|I_N)\}$ now available to \mathcal{D} . Formal subjective Bayesian analysis dictates that \mathcal{D} will then use the information set I to predict y_t using the implied posterior $p(y_t|I)$ from a full Bayesian prior-to-posterior analysis. By extending the extant theory in Genest and Schervish (1985); West and Crosse (1992), West (1992) showed that there is a subset of all Bayesian models in which \mathcal{D} 's posterior has the mathematical form

$$p(y_t|I) = \int_{\tilde{y}_t} \alpha(y_t|\tilde{y}_t) \prod_{i=1}^N p(\tilde{y}_{it}|I_i) d\tilde{y}_t \quad (17)$$

where, to use our earlier notation, \tilde{y}_t is now a latent N -dimensional vector comprising the latent \tilde{y}_{it} ; and $\alpha(y_t|\tilde{y}_t)$ is a conditional p.d.f. for y_t given \tilde{y}_t . For more details on interpretation we refer to West (1992).

There are two central aspects within the BPS framework which are important to clarify in relation to earlier combination methods in economics.

Firstly, the synthesis function, $\alpha(y_t|\tilde{y}_t)$, is essentially the fundamental combination equation defined in the previous section i.e. $p(y_t|\tilde{y}_t)$. As stated there, the real issue is how to specify and define its functional form. McAlinn and West (2018) show that many forecast and model combination methods including those of the linear opinion pool, (2), and the density forecast combinations considered in Geweke and Amisano (2011); Kapetanios et al. (2015); Pettenuzzo and Ravazzolo (2016); Aastveit et al. (2018) can be considered as special cases of (17), realized via different choices of the form of the BPS

⁵For extensions of BPS to account for agent/model inter-dependencies or to a multivariate setting see Johnson and West (2017) and McAlinn et al. (2017), respectively.

synthesis function $\alpha(y_t|\tilde{y}_t)$.

Secondly, is the interpretation of the latent vector \tilde{y}_t . Suppose the agent provides a density degenerate at a point, x_{it} i.e. $p(y_t|I_i) = \delta_{x_{it}}(y_t)$ for $i = 1, \dots, N$. That is, \mathcal{A}_i makes a perfect prediction $y_t = x_{it}$ for some specified value x_{it} . \mathcal{D} 's posterior is then $\alpha(y_t|\tilde{y}_t)$ which reflects the views of y_t based on supposedly exact predicted values from the agents. Thus, one can refer to x_{it} as the latent agent states and to $\alpha(y_t|\tilde{y}_t)$ as \mathcal{D} 's calibration function.

The key issue is defining the form or model for $\alpha(y_t|\tilde{y}_t)$. McAlinn and West (2018) consider an example where the calibration density can be rewritten as

$$\alpha(y_t|\tilde{y}_t) = N(y_t|F_t'\theta, v) \quad (18)$$

with $F = (1, \tilde{y}_t)'$ and $\theta = (\theta_0, \theta_1, \dots, \theta_N)'$. This means that the practically relevant effective calibration parameters are (θ, v) , and historical data will inform \mathcal{D} on these. Next, to illustrate further the approach, we follow McAlinn and West (2018) and consider the extension of the BPS framework to sequential forecasting of time series.

Dynamic Sequential Setting Consider that the decision maker \mathcal{D} receives forecast densities from each agent sequentially over time. At time $t - 1$, \mathcal{D} receives current forecast densities $I_t = \{p(y_t|I_1), \dots, p(y_t|I_N)\}$ from the set of agents and aims to forecast y_t . The full information set used by \mathcal{D} at time t is thus $\{y_{1:t-1}, I_{1:t}\}$. As \mathcal{D} observes more information, her views of the agent biases and calibration characteristics, as well as of inter-dependencies among agents are repeatedly updated. A formal, parameterized Bayesian dynamic model is the vehicle for structuring this sequential learning in a general state-space context. This defines the dynamic BPS framework.

Consider again the dynamic regression for the BPS synthesis function

$$\alpha(y_t|\tilde{y}_t, \Phi_t) = N(y_t|F_t\theta_t, v_t) \quad (19)$$

where $F_t = (1, \tilde{y}_t)'$ and $\theta_t = (\theta_{t0}, \theta_{t1}, \dots, \theta_{tN})'$ being a $1 + N$ -vector of time-varying bias/calibration coefficients (weights).

Modelling time evolution of the parameter processes $\Phi_t = (\theta_t, v_t)$ is needed to complete model specification. McAlinn and West (2018) do this using the first step into dynamic models, with traditional random walk models to allow for– but not anticipate direction in– stochastic changes over time in both regressions θ_t and volatilities v_t , as is traditional in Bayesian time series literatures; see, for example, West and Harrison (1997) (chap. 16) and Prado and West (2010) (chap. 10). Thus McAlinn and West (2018) take

$$y_t = F_t \theta_t + \nu_t, \quad \nu_t \sim N(0, v_t), \quad (20a)$$

$$\theta_t = \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, v_t W_t) \quad (20b)$$

where θ_t evolves in time according to a linear/normal random walk with innovations variance matrix $v_t W_t$ at time t , and v_t is the residual variance in predicting y_t based on past information and the set of agent forecast distributions.

We summarize the analysis of this and the previous section as follows. The fundamental equation of section 3.4.1, equation (11), relates to the BPS equation (17) of the present section as follows. The BPS approach provides a formal Bayesian decision based framework; and makes this operational by specifying probability distributions like the normal one that are easy to use for numerical evaluation of the forecast density combination approach. The flexible density combination approach of section 3.4.1 and Billio et al. (2013) is a very general specification of forecast density combinations. It allows for probability distributions that are not members of known classes of distribution functions and it uses probabilistic weights in the density combinations that constitute a convex set in the unit interval. As a consequence, this approach makes use of a representation result which states that the forecast density combination model can be represented as a nonlinear, non-Gaussian state space model where filtering methods are used for the numerical

evaluation of the distributions involved. If one assumes more restricted versions of the distributions in the density combination approach, then one can avoid particle filtering methods and one can make use of standard MCMC procedures.

3.5 Comments

Inferential procedures We emphasize that so far we have not explicitly specified the details of the method of inference. We suggest a Bayesian approach using simulation methods. The attractiveness of the Bayesian approach is that one may interpret the predictive densities of the different models as *prior* densities and the conditional density of y_t as a *likelihood*. This shows that the learning aspect, which is typical for a standard Bayesian analysis, also extends to the density combination approach.

However, given the general specification of the density combinations model, one may also make use of maximum likelihood inference or Generalized Methods of Moments (GMM). The results are then only approximately exact, depending on the size of the sample, while the simulation based Bayesian results hold for the finite sample.

Model set choice and incompleteness Model set choice and incompleteness of this set has increasingly become an issue in the economics literature. We present a brief summary of the evolution of this literature. Hall and Mitchell (2007) considered combining (just) two or three density forecasts, from competing macroeconomic forecasters. But, as inspection of (2) reveals, as N increases, in principle, the combined density forecast becomes more and more flexible and better able to approximate non-linear and non-Gaussian data-generating-processes. Thus an open issue is whether it is preferable to combine a small or large number of forecast densities. In many applications, the choice of N may be natural, indeed fixed, and there is nothing further to discuss in these cases. But in other applications there may be discretion about what models to include and how many. Indeed this is the situation in many economic applications, where given fears of

model incompleteness, it often appears attractive to entertain many different forecast densities distinguished, perhaps, by the information set on which they condition. As recent datasets in economics often have a large cross-sectional dimension, it is easy to think of a large number of models differentiated, for example, by what economic variables they condition the forecast on. Jore et al. (2010) combine a large number of forecast densities from models differentiated according to how, amongst other things, they model suspected structural breaks and instabilities in their macroeconomic sample; to remain agnostic about when the structural changes occur, they entertain breaks at any point in the sample giving rise to a large number of potentially similar forecasting models.

By contrast, Aastveit et al. (2014), for example, chose to adopt a two-step approach that firstly combines the forecast densities from models within a specific model class, and then combines these combined forecast densities across the smaller number of model classes. While Mazzi et al. (2014) consider dropping *bad* models from their linear combinations. These different strategies are, in effect, different ways of accommodating the dependence that no doubt exists between the competing forecast densities; cognizant of earlier work that showed dependence among forecasts greatly reduces the gains in performance from increasing the number of forecasts being combined (e.g. see Clemen and Winkler (1985)).

4 Selected Applications

Our purpose in this section is to illustrate the recent history and evolution of, as well as the potential and benefits of, forecast density combinations in policy making environments by reviewing a few selected applications.

4.1 Combining macroeconomic forecasts at central banks

Forecasts are now central to policy making and communication at many central banks, given that monetary policy works with a lag. Many central banks, including the Norges Bank discussed further below, publish variants of the density forecast or fan chart seen above in Figure 1. While these forecasts, typically published by the central bank in some form of inflation report, may involve the use of both econometric models and judgement, it has become increasingly widespread for central banks to consult a range of econometric models when producing their model-based density forecasts. This reflects the “model uncertainty” as evidenced by the wide range of econometric and structural models being regularly consulted and used by central banks. These models include single equation regression models, univariate models, mixed-frequency models and multivariate models such as VAR models, dynamic factor models and structural DSGE models; e.g. see Alessi et al. (2014). In the face of this model uncertainty, the forecast density combination methods considered in this chapter have become popular, indeed many have been developed at or in partnership with central banks.

Applications by Mitchell and Hall (2005) and Hall and Mitchell (2007) indicated that the density forecasts of inflation from the Bank of England (as published in their fan charts) and the National Institute of Economic and Social Research in the UK could be improved if combined in real-time with some simple time-series forecasts using optimized (with respect to the log score) weights. However, equal weighted forecast density combinations fared, in their applications, worse than optimized weights or indeed weights that, in effect, select a single (best) model. Subsequent research, much but not all with a central banking focus and including Jore et al. (2010), Bache et al. (2011) and Bache et al. (2010), extended the forecast density combinations to consider a large number of models. These ‘models’ are differentiated not simply according to their model ‘class’ (e.g. VAR, dynamic factor model or DSGE model) but according to additional modeling assumptions made

for a given ‘type’ of model. For example, in an application to US output growth, inflation and interest rates, Jore et al. (2010) find that forecast density combinations from a large number of VAR models work well if the combination weights favor those VAR models that allow for structural breaks. In another application to US output growth and inflation, but consulting a larger number of macroeconomic predictors, Rossi and Sekhposyan (2014) find that equal and Bayesian weighted forecast density combination work well. Amisano and Geweke (2017) also find that equal weighted linear combinations of forecast densities from three models commonly used at central banks (namely, dynamic factor, DSGE and VAR models) work well in an application to quarterly US data. And Del Negro et al. (2016) find strong evidence of time variation in the combination weights reflecting the fact that a DSGE model with financial frictions produces superior forecasts, to a DSGE model without financial frictions, in periods of financial distress but does not perform as well in tranquil periods.

In short, while the forecast density combinations have been tuned to the application of interest, they have been found to be an effective means of producing better calibrated and more robust macroeconomic forecasts in central banking contexts.

4.1.1 SAM: at the Norges Bank

To provide an explicit example of the use of forecast density combination methods, in this section we consider the System of Averaging Models (SAM) as operationalized at the Norges Bank. SAM involves combining a large set of models in order to produce not only point forecasts but also density forecasts for the Norwegian economy. Since 2008 SAM has been used to provide the Bank with model-based short-term density forecasts (up to five quarters ahead) for Mainland GDP and consumer prices adjusted for taxes and without energy (CPIATE). The forecasts produced by SAM are combinations of density forecasts for quarterly growth in Mainland GDP and four-quarter growth in CPIATE, on the basis of the flow of information that becomes available during the quarter. While internally the

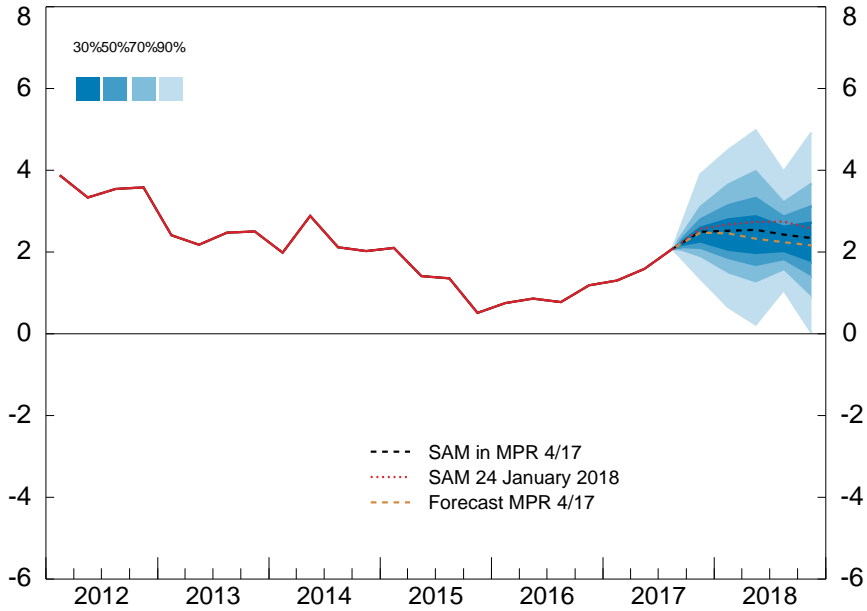
combined density forecasts are regularly updated (in principle they can be updated every day there is a new data release), the SAM forecasts are made publicly available at Norges Bank's internet page in conjunction with each monetary policy meeting in Norges Bank's Executive Board.⁶ The SAM forecasts are also published in every *Monetary Policy Report* (MPR) together with Norges Bank's official forecasts. SAM provides pure model-based forecasts, while Norges Bank's final "official" short-term forecasts are, in general, subject to judgement.

Figure 2 depicts the fan charts for GDP and CPIATE from SAM published on Norges Bank's web site on 24 January 2018. When MPR 4/17 was published, in December 2017, GDP was judged to increase somewhat less than the mean of the SAM densities, while inflation was judged to be higher than the mean of the SAM densities. At the monetary policy meeting in January, SAM forecasts for both growth GDP and inflation were revised upwards.

The density forecasts in SAM are combined in a two-step procedure, as briefly mentioned in section 3.5 above. In the first step models are grouped into three different model classes: VAR models, factor models and leading indicator (bridge equation) models. Forecasts are then combined within each model class, using their logarithmic scores (log score) to compute their weights (see, among others, Jore et al. (2010)). This yields a combined density forecast for each of the three model classes. In a second step, these three predictive densities are combined into a single density nowcast, again using log score weights. The advantages of this approach are that it explicitly accounts for uncertainty of model specification and instabilities within each model class, and that it implicitly gives *a priori* equal weight to each model class. For a more detailed description of SAM see Aastveit et al. (2011); see also Aastveit et al. (2014) who document the usefulness of this approach for nowcasting US GDP growth.

⁶The most recent SAM forecasts for the Norwegian economy are available here: <https://www.norges-bank.no/en/Monetary-policy/Models-for-monetary-policy-analysis-and-forecasting/SAM/>.

GDP



CPI-ATE

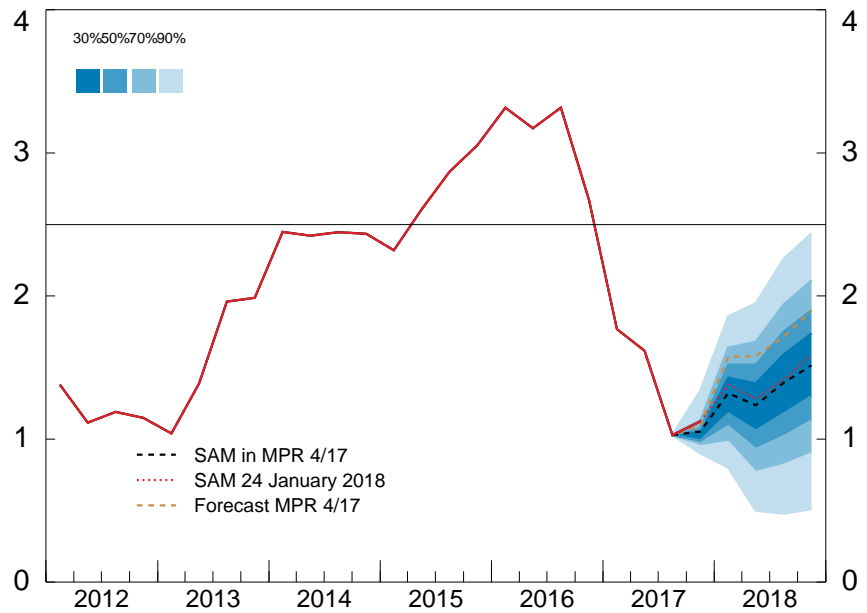


Figure 2: Density forecasts for Mainland GDP and CPI-ATE from SAM and point forecasts from MPR 4/17. Four-quarter growth. Percent. Source: Norges Bank

5 Remarks on Further Evolution

This paper provides a focused survey on the evolution of forecast density combination methods in economics; it organizes its discussion around the *fundamental density combination equation* and discusses how various frequentist as well as Bayesian approaches can be seen as giving it different specific content. The paper explains how recent evolutions of forecast density combination methods have sought to specify more flexible combination densities that allow for time-variation, learning features and model incompleteness. Evolution continues.

We anticipate that future evolutions may involve:

- Forecast density combinations with “big data”. Increasingly large often micro-based datasets and surveys are available. This is especially so in countries with advanced data collection methods, like Sweden and Norway for example, where information is increasingly becoming available not just at a macroeconomic (or aggregated) but at a micro (individual or disaggregated) level. So nowcasting and forecasting methods should aim to harness these “population” often online data, rather than representative samples of data as typically captured, for example, by national accounts data; and benefit from a broader, higher-frequency and faster-moving information set.
- Exploiting improvements in computer software and hardware, including advanced parallel computing. With the arrival of “big data”, but with in effect “model uncertainty” since it is often unclear how to model and forecast with such a huge number of potential *indicator* variables, forecast density combination methods remain a natural tool. Methods will evolve by drawing on and adapting machine learning methods, like ‘deep neural nets’, to discover and exploit data patterns without much *a priori* theory. One interesting development is to combine micro and macroeconomic data with more detailed financial, micro-based data. Forecast density combinations

will need increasing flexibility to accommodate the shape of the “data” distribution, to reflect perhaps severe departures from Gaussianity for data distributions that are highly skewed, with multi-modalities and fat tails or outliers; this presents new analytical and computational challenges for the implementation of forecast density combinations.

- Increasing attention being paid to the production, communication and use of multi-variate forecast densities, that seek to model and represent the dependence between the marginal density forecasts. As an example, re-consider the Bank of England’s fan charts. As discussed, these are presented separately for GDP growth, inflation and unemployment. But policy makers are naturally interested in the dependencies between these three forecasts, and what these dependencies mean for policy. And density forecast combination methods should also evolve to let decision and policy makers condition these forecasts on the broader, higher often mixed-frequency and faster-moving information sets that increasingly are available.

Therefore, we anticipate many further evolutions of forecast density combination methods and applications in the field of economic forecasting. However, we end this paper on a well-known note of caution that also serves as an encouragement, we hope, for future research. Major challenges remain of how best to link (density) forecasts to decisions actually made (using the forecasts) so that, for example, the effects of alternative economic policies or decisions can be identified and evaluated. This suggests that to aid decision and policy making in practice there should be a closer link between structural economic models, decision theory and how density forecast combinations are produced, used and evaluated.

Further Readings

Increasingly, professional forecasters and academic researchers present model-based and subjective or judgment-based forecasts in economics which are accompanied by some measure of uncertainty. In its most complete form this measure is a probability density function for future values of the variables of interest. How to combine forecast densities from several sources like experts or models is therefore a very active research area within economics and statistics. Previous reviews and textbook treatments of forecast density combinations in economics are given by Hall and Mitchell (2009) and Elliott and Timmermann (2016). For further readings, more recent developments range from combining predictive densities using weighted linear combinations of prediction models, evaluated using various scoring rules, see Hall and Mitchell (2007), Geweke and Amisano (2011), Jore et al. (2010) and Aastveit et al. (2014), to more complex combination approaches that allows for time-varying weights with possibly both learning and model set incompleteness, see Billio et al. (2013), Casarin et al. (2015), Del Negro et al. (2016), Aastveit et al. (2018) and McAlinn et al. (2017). For multivariate combinations and combinations with many predictors, see and McAlinn and West (2018) and Casarin et al. (2018), respectively.

References

- Aastveit, K., K. Gerdrup, and A. Jore (2011). Short-term forecasting of GDP and inflation in real time: Norges bank's system for averaging models. Staff Memo 2011/9, Norges Bank.
- Aastveit, K., K. Gerdrup, A. Jore, and L. Thorsrud (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business and Economic Statistics* 32(1), 48–68.
- Aastveit, K. A., F. Ravazzolo, and H. K. van Dijk (2018). Combined density Nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics* 36(1), 131–145.
- Ager, P., M. Kappler, and S. Osterloh (2009). The accuracy and efficiency of the consensus forecasts: A further application and extension of the pooled approach. *International Journal of Forecasting* 25(1), 167 – 181.
- Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014). Central bank macroeconomic forecasting during the global financial crisis: The european central bank and federal reserve bank of new york experiences. *Journal of Business & Economic Statistics* 32(4), 483–500.
- Amisano, G. and J. Geweke (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216–230.
- Amisano, G. and J. Geweke (2017). Prediction using several macroeconomic models. *The Review of Economics and Statistics* 99(5), 912–925.
- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. 25(2), 177–190.
- Bache, I. W., A. S. Jore, J. Mitchell, and S. P. Vahey (2011). Combining VAR and DSGE forecast densities. *Journal of Economic Dynamics and Control* 35(10), 1659–1670.
- Bache, I. W., J. Mitchell, F. Ravazzolo, and S. P. Vahey (2010). Macro modelling with many models. In D. Cobham, y. Eitrheim, S. Gerlach, and J. F. Qvigstad (Eds.), *Twenty years of inflation targeting: lessons learnt and future prospects*, pp. 398–418. Cambridge University Press.
- Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance* 55, 225–264.
- Bassetti, F., R. Casarin, and F. Ravazzolo (2018). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*. Forthcoming.

- Bastürk, N., A. Borowska, S. Grassi, L. Hoogerheide, and H. K. Van Dijk (2018). Time-varying learning combinations of Bayesian dynamic models and equity momentum strategies. Forthcoming.
- Bastürk, N., L. Hoogerheide, and H. K. van Dijk (2017). Bayesian analysis of boundary and near-boundary evidence in econometric models with reduced rank. *Bayesian Analysis* 12(3), 879–917.
- Batchelor, R. and P. Dua (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science* 41, 68–75.
- Bates, J. and C. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20(4), 451–468.
- Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics* 177, 213–232.
- Busetti, F. (2017). Quantile aggregation of density forecasts. *Oxford Bulletin of Economics and Statistics* 79(4), 495–512.
- Casarin, R., S. Grassi, F. Ravazzolo, and H. K. van Dijk (2015). Parallel sequential monte carlo for efficient density combination: The deco matlab toolbox. *Journal of Statistical Software* 68(3).
- Casarin, R., S. Grassi, F. Ravazzolo, and H. K. van Dijk (2018). Predictive density combinations with dynamic learning for large data sets in economics and finance. Technical report.
- Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics* 29(3), 327–341.
- Clark, T. E. and M. W. McCracken (2010). Averaging forecasts from vars with uncertain instabilities. *Journal of Applied Econometrics* 25(1), 5–29.
- Clemen, R. and R. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* 19, 187–203.
- Clemen, R. T. and R. L. Winkler (1985). Limits for the precision and value of information from dependent sources. *Operations Research* 33(2), 427–442.
- Clements, M. P. (2002). Comments on 'The state of macroeconomic forecasting'. *Journal of Macroeconomics* 24(4), 469–482.
- Clements, M. P. (2004). Evaluating the Bank of England density forecasts of inflation. *Economic Journal* 114.
- Conflitti, C., C. De Mol, and D. Giannone (2015). Optimal combination of survey forecasts. *International Journal of Forecasting* 31(4), 1096–1103.

- Corradi, V. and N. R. Swanson (2006). Predictive density evaluation. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 197–284. North Holland: North-Holland.
- Del Negro, M., B. R. Hasegawa, and F. Schorfheide (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics* 192(2), 391–405.
- Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39(4), 863–83.
- Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1), 45–97.
- Elliott, G. (2017). Forecast combination when outcomes are difficult to predict. *Empirical Economics* 53(1), 7–20.
- Elliott, G. and A. Timmermann (2016). *Economic Forecasting*. Princeton University Press.
- Fernandez-Villaverde, J. and J. Rubio-Ramirez (2004). Comparing dynamic equilibrium economies to data: A Bayesian approach. *Journal of Econometrics* 123, 153–187.
- Galvao, A., A. Garratt, and J. Mitchell (2018). Comparing alternative methods of combining density forecasts - with an application to us inflation and gdp growth.
- Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin (2003). Forecast uncertainties in macroeconomic modeling: An application to the UK economy. *Journal of the American Statistical Association* 98(464), 829–38.
- Garratt, A., J. Mitchell, S. P. Vahey, and E. C. Wakerly (2011). Real-time inflation forecast densities from ensemble Phillips curves. *The North American Journal of Economics and Finance* 22(1), 77–87.
- Genest, C. and M. J. Schervish (1985). Modelling expert judgements for Bayesian updating. *Annals of Statistics* 13, 1198–1212.
- Genest, C. and J. V. Zidek (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science* 1(1), 114–148.
- Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics* 164(1), 130 – 141.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* 106, 746–762.

- Gneiting, T., A. E. Raftery, A. H. W. III, and T. Goldman (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review* 133(5), 1098–1118.
- Gneiting, T. and R. Ranjan (2013). Combining predictive distributions. *Electron. J. Statist.* 7, 1747–1782.
- Granger, C. and M. Pesaran (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19, 537–560.
- Granger, C. W. J. and R. Ramanathan (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting* 3, 197–204.
- Granger, C. W. J., H. White, and M. Kamstra (1989). Interval forecasting : An analysis based upon ARCH-quantile estimators. *Journal of Econometrics* 40(1), 87–96.
- Grushka-Cockayne, Y., V. R. R. Jose, and K. C. Lichtendahl Jr. (2017). Ensembles of overfit and overconfident forecasts. *Management Science* 63(4), 1110–1130.
- Guidolin, M. and A. Timmermann (2009). Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach. *Journal of Econometrics* 150, 297–311.
- Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting* 23(1), 1–13.
- Hall, S. G. and J. Mitchell (2009). Recent developments in density forecasting. In T. C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, pp. 199–239. Palgrave MacMillan.
- Hoogerheide, L., R. Kleijn, R. Ravazzolo, H. K. van Dijk, and M. Verbeek (2010). Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights. *Journal of Forecasting* 29(1-2), 251–269.
- Hora, S. C. (2004). Probability judgments for continuous quantities: Linear combinations and calibration. *Management Science* 50(5), 597–604.
- Hora, S. C., B. R. Fransen, N. Hawkins, and I. Susel (2013). Median aggregation of distribution functions. *Decision Analysis* 10(4), 279–291.
- Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton (1991). Adaptive mixtures of local experts. *Journal of Neural Computation* 3, 79–87.
- Johnson, M. C. and M. West (2017). Bayesian predictive synthesis for probabilistic forecast calibration and combination.
- Jordan, M. and R. Jacobs (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation* 6, 181–214.

- Jordan, M. I. and L. Xu (1995). Convergence results for the EM approach to mixtures of experts architectures. *Neural Networks* 8, 1409–1431.
- Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining forecast densities from vars with uncertain instabilities. *Journal of Applied Econometrics* 25(4), 621–634.
- Jouini, M. N. and R. T. Clemen (1996). Copula models for aggregating expert opinions. *Operations Research* 44, 444–457.
- Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett (2015). Generalised density forecast combinations. *Journal of Econometrics* 188, 150–165.
- Kascha, C. and F. Ravazzolo (2010). Combining inflation density forecasts. *Journal of Forecasting* 29(1-2), 231–250.
- Knüppel, M. (2015). Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments. *Journal of Business & Economic Statistics* 33(2), 270–281.
- Knüppel, M. and G. Schulte-frankenfeld (2012). How Informative Are Central Bank Assessments of Macroeconomic Risks? *International Journal of Central Banking* 8(3), 87–139.
- Krüger, F. (2017). Survey-based forecast distributions for euro area growth and inflation: ensembles versus histograms. *Empirical Economics* 53(1), 235–246.
- Lichtendahl Jr., K. C., Y. Grushka-Cockayne, and R. L. Winkler (2013). Is it better to average probabilities or quantiles? *Management Science* 59(7), 1594–1611.
- Lutz, K. and S. Manganelli (2007). Quantifying the risk of deflation. *Journal of Money, Credit and Banking* 39(2-3), 561–590.
- Maheu, J. M. and T. H. McCurdy (2009). How useful are historical data for forecasting the long-run equity return distribution? *Journal of Business & Economic Statistics* 27(1), 95–112.
- Mazzi, G., J. Mitchell, and G. Montana (2014). Density nowcasts and model combination: nowcasting euro-area gdp growth over the 2008-9 recession. *Oxford Bulletin of Economics and Statistics* 76(2), 233–256.
- McAlinn, K., K. A. Aastveit, J. Nakajima, and M. West (2017). Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Mimeo*.
- McAlinn, K. and M. West (2018). Dynamic bayesian predictive synthesis for time series forecasting. *Journal of Econometrics*. Forthcoming.
- Mitchell, J. (2013). The Recalibrated and Copula Opinion Pools. EMF Research Papers 02, Economic Modelling and Forecasting Group.

- Mitchell, J. and S. G. Hall (2005). Evaluating, comparing and combining density forecasts using the klic with an application to the bank of england and niersr 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics* 67(s1), 995–1033.
- Mitchell, J. and K. F. Wallis (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics* 26, 1023–1040.
- Morris, P. (1974). Decision analysis expert use. *Management Science* 20, 1233–1241.
- Morris, P. (1977). Combining expert judgments: A Bayesian approach. *Management Science* 23, 679–693.
- Opschoor, A., D. van Dijk, and M. van der Wel (2017). Combining density forecasts using focused scoring rules. *Journal of Applied Econometrics* 32(7), 1298–1313.
- Pauwels, L. L. and A. L. Vasnev (2016). A note on the estimation of optimal weights for density forecast combinations. *International Journal of Forecasting* 32(2), 391–397.
- Peng, F., R. A. Jacobs, and M. A. Tanner (1996). Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. *Journal of the American Statistical Association* 91, 953–960.
- Pettenuzzo, D. and F. Ravazzolo (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics* 31(7), 1312.
- Prado, R. and M. West (2010). *Time Series: Modelling, Computation & Inference*. Chapman & Hall/CRC Press.
- Raftery, A., M. Karny, and P. Ettlter (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics* 52, 52–66.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review* 133, 1155–1174.
- Ranjan, R. and T. Gneiting (2010). Combining probability forecasts. *Journal of the Royal Statistical Society Series B* 72(1), 71–91.
- Roberts, H. V. (1965). Probabilistic prediction. *Journal of American Statistical Association* 60, 50–62.
- Rossi, B. (2013). Advances in Forecasting under Instabilities. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2*. Elsevier-North Holland Publications.

- Rossi, B. and T. Sekhposyan (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting* 30(3), 662–682.
- Sanders, F. (1963). On subjective probability forecasting. *Journal of Applied Meteorology (1962-1982)* 2(2), 191–201.
- Sloughter, J., T. Gneiting, and A. E. Raftery (2010). Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging. *Journal of the American Statistical Association* 105, 25–35.
- Smith, J. and K. F. Wallis (2009). A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics* 71(3), 331–355.
- Stock, J. H. and M. W. Watson (2004). Combining forecasts of output growth in seven-country data set. *Journal of Forecasting* 23, 405–430.
- Terui, N. and H. K. van Dijk (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting* 18, 421–438.
- Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 136–96. Amsterdam: Elsevier.
- Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics* 67(s1), 983–994.
- Wallis, K. F. (2011). Combining forecasts - forty years later. *Applied Financial Economics* 21, 33–41.
- West, M. (1992). Modelling agent forecast distributions. *Journal of the Royal Statistical Society (Series B: Methodological)* 54, 553–567.
- West, M. and J. Crosse (1992). Modelling of probabilistic agent opinion. *Journal of the Royal Statistical Society (Series B: Methodological)* 54, 285–299.
- West, M. and P. J. Harrison (1997). *Bayesian Forecasting & Dynamic Models* (2nd ed.). Springer Verlag.
- Winkler, R. (1968). The consensus of subjective probability distributions. *Management Science* 15, B61–B75.
- Winkler, R. (1981). Combining probability distributions from dependent information sources. *Management Science* 27, 479–488.