

TI 2011-082/4  
Tinbergen Institute Discussion Paper



# Bayesian Combinations of Stock Price Predictions with an Application to the Amsterdam Exchange Index

*Monica Billio<sup>1</sup>*

*Roberto Casarin<sup>1</sup>*

*Francesco Ravazzolo<sup>2</sup>*

*Herman K. van Dijk<sup>3</sup>*

<sup>1</sup> *University of Venice, GRETA Ass. and School for Advanced Studies in Venice;*

<sup>2</sup> *Norges Bank;*

<sup>3</sup> *Econometrics and Tinbergen Institutes, and Erasmus University Rotterdam.*

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

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)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Bayesian Combinations of Stock Price Predictions with an Application to the Amsterdam Exchange

## Index\*

Monica Billio<sup>†</sup>      Roberto Casarin<sup>†</sup>

Francesco Ravazzolo<sup>§</sup>      Herman K. van Dijk<sup>‡\*\*</sup>

<sup>†</sup>University of Venice, GRETA Assoc. and School for Advanced Studies in Venice

<sup>§</sup>Norges Bank

<sup>‡</sup>Econometrics and Tinbergen Institutes and Erasmus University Rotterdam

May 2, 2011

### Abstract

This paper deals with the problem of combining predictive densities for financial series. We summarize the general combination approach based on a Bayesian state space representation of the predictive densities and of the combination scheme which allows for incomplete model space proposed by Billio et al. [2010]. In the combination model the weights follow logistic autoregressive processes, change over time and their dynamics are possibly driven by the past forecasting performances of the predictive densities. For illustrative purposes we apply it to combine White Noise and GARCH models to forecast the Amsterdam Exchange index and use the combined predictive forecasts in an investment asset allocation exercise.

*JEL codes:* C11, C15, C53, E37.

*Keywords:* Density Forecast Combination, Stock data.

---

\*The views expressed in this paper are our own and do not necessarily reflect those of Norges Bank.

\*\*Corresponding author: hkvandijk@ese.eur.nl

# 1 Introduction

When multiple forecasts are available from different models or sources it is possible to combine them in order to make use of all the available information on the variable to be predicted and, as a consequence, to possibly produce better forecasts. Most of applications concern macroeconomic data, but less evidence there is with financial data. This paper deals with the problem of combining predictive densities for financial series. Following Billio et al. [2010], we apply a general combination approach based on a Bayesian state space representation of the predictive densities and of the combination scheme. In the combination model the weights follow logistic autoregressive processes change over time, and their dynamics can be driven by the past forecasting performances of the predictive densities. Moreover, parameter, model and weight uncertainty is taken into account in the combination scheme. For illustrative purposes we provide an application to forecast and allocate a portfolio in the Amsterdam Exchange Stock Market.

In the literature there is growing interest in model combination and many different approaches have been proposed. Bayesian Model Averaging (BMA) is one of the most common procedure, see Hoeting et al. [1999] for a review on BMA, with an historical perspective. BMA gives a probability to any individual models and combines them to obtain point and density forecasts. Parameter uncertainty and model uncertainty are taken into account, but the model space is assumed to be complete and the correct model is supposed to exist (in the limit). Our work builds on another stream of literature started with Bates and Granger [1969] and combines predictions from different forecasting models. See Granger [2006] for an updated review on forecast combination. Granger and Ramanathan [1984] extend Bates and Granger [1969] and propose to combine forecasts with weights obtained as unrestricted regression coefficients. Terui and van Dijk [2002] generalize the least squares weights by representing the dynamic forecast combination as a state space

model. In their work the weights are assumed to follow random walk processes. Hoogerheide et al. [2010] and Groen et al. [2009] propose robust time-varying weights. Recently Billio et al. [2010] extend the state-space representation of Terui and van Dijk [2002] and Hoogerheide et al. [2010] by assuming time-varying  $[0,1]$  (e.g. logistic transformation) weights and propose a Bayesian state-space representation of the predictive densities and of the combination scheme. In these papers the model space is possibly incomplete.

Predictability for financial returns is very low and all models seem wrong and unable to capture all the dynamics in such markets; and their performance varies substantially over time. These features should be considered when combining financial return predictive densities, making standard combination schemes unappealing. Following Billio et al. [2010], we represent our combination schemes in terms of conditional densities and write equations for producing predictive densities and not point forecasts (as is often the case) for the variables of interest. We consider convex combinations of the predictive densities and assume that the time-varying weights associated with the different predictive densities belong to the standard simplex. Under this constraint the weights can be interpreted as a discrete probability distribution over the set of predictors. Tests for a specific hypothesis on the values of the weights can be conducted due to their random nature. The weighting schemes have time-varying continuous dynamics and a learning mechanism is also introduced to allow the dynamics of each weight to be driven by the past and current performances of the predictive densities in the combination scheme. A Sequential Monte Carlo (SMC) algorithm similar to Billio et al. [2010] is applied.

The application shows that the methodology improves statistical accuracy of forecasts for the Amsterdam Stock Exchange index, in particular in terms of density forecasting. However, this does not guarantee higher investment performance.

The structure of the paper is as follows. Section 2 describes the combinations

of prediction densities and presents time-varying weights dynamics with learning mechanism. Section 3 provides the results of the application of the proposed combination method to the macroeconomic and financial datasets. Section 4 concludes.

## 2 Bayesian Densities Combinations

Let  $y_t$  be an observable variable at time  $t$ , with  $t = 1, \dots, T$ : we are interested in predicting the future values of the variable  $y_t$ . In particular, in a density forecasting exercise, we are interested in estimating  $p(y_t|y_{1:t-1})$ , which is the distribution of the  $y_t$  conditional on its past values and which is called one-step-ahead prediction density of  $y_t$ . In many situations there are different prediction models available for the variable  $y_t$ . In what follows we will assume that at time  $t$  a set of  $K$  one-step-ahead predictors  $\tilde{y}_{k,t}$ , with  $k = 1, \dots, K$ , is available from different models or sources. Moreover we assume that for each prediction model its conditional density  $p(\tilde{y}_{k,t}|y_{1:t-1})$  is available analytically or in an approximated form (e.g. through Monte Carlo samples).

When many prediction models are available one of the challenging issues is to summarize the information on the future values of the variable. The combination of predictions represents a solution to this problem. We propose here an optimal combination based on the distributional representation of the predictive models and, following Billio et al. [2010], we suggest to summarize the information from the different predictive densities in one prediction density for  $y_t$  by conditioning on  $\tilde{\mathbf{y}}_t = (\tilde{y}_{1,t}, \dots, \tilde{y}_{K,t})$  and on a combination scheme  $\mathbf{w}_t = (w_{1,t}, \dots, w_{K,t})$

$$p(y_t|\mathbf{w}_t, \tilde{\mathbf{y}}_t) \propto \exp \left\{ -\frac{1}{2\sigma^2} (y_t - \mathbf{w}'_t \tilde{\mathbf{y}}_t)^2 \right\} \quad (1)$$

which corresponds to a Gaussian combination where  $\mathbf{w}_t$  are the weights:

$$w_{k,t} = \frac{\exp\{x_{k,t}\}}{1 + \sum_{j=1}^{K-1} \exp\{x_{j,t}\}}, \quad \text{with } k = 1, \dots, K-1 \quad (2)$$

$$w_{K,t} = 1 - \sum_{j=1}^K \exp\{x_{K,t}\} \quad (3)$$

The weights are thus multivariate logistic transformations of a latent process  $\mathbf{x}_t$ . The transformation allows for positive weights that sum to one and accordingly can be interpreted as the probability associated to a specific prediction model. In this work we assume that the latent factor has the following Gaussian dynamics

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, \tilde{\mathbf{y}}_{1:t-1}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{x}_t - \mathbf{x}_{t-1} + \Delta \mathbf{e}_t)' \Lambda^{-1} (\mathbf{x}_t - \mathbf{x}_{t-1} + \Delta \mathbf{e}_t) \right\} \quad (4)$$

with exogenous variable  $\Delta \mathbf{e}_t = \mathbf{e}_t - \mathbf{e}_{t-1}$ , where  $\mathbf{e}_t = (e_{1,t}, \dots, e_{K,t})$  is a vector of exponentially weighted average errors

$$e_{k,t} = (1 - \lambda) \sum_{i=1}^{\tau} \lambda^{i-1} (y_{t-i} - \hat{y}_{k,t-i})^2 \quad (5)$$

with  $\lambda \in (0, 1)$  being a smoothing parameter and  $\tau$  the size of the window of evaluation of past errors. The past forecasting performance of the predictors is thus included in the weights dynamics. A deterioration of the forecasting performance of the  $k$ -th prediction model (i.e.  $\Delta e_{k,t} > 0$ ) reduces its weight in the combination (i.e.  $w_{k,t}$  decreases). As opposite, an improvement in the prediction performance (i.e.  $\Delta e_{k,t} < 0$ ) increases the value of the  $k$ -th weight. This simple mechanism has been originally proposed Diebold and Pauly [1987] without the random components.

This combination scheme represents a general relationship between observable, model-specific predictive densities, combination weights and the predictive density

for  $y_t$

$$p(y_t|\mathbf{y}_{1:t-1}) = \int \int p(y_t|\mathbf{w}_t, \tilde{\mathbf{y}}_t) p(\mathbf{w}_t|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) p(\tilde{\mathbf{y}}_{1:t}|\mathbf{y}_{1:t-1}) d\mathbf{w}_t d\tilde{\mathbf{y}}_{1:t} \quad (6)$$

This relationship for the prediction of the observable variable  $y_t$  is part of a general filtering and prediction problem which can be represented conditionally on  $\tilde{\mathbf{y}}_{1:t}$  through the following set of recursions

$$p(\mathbf{w}_t|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t}) \propto p(y_t|\mathbf{w}_t, \tilde{\mathbf{y}}_t) p(\mathbf{w}_t|\mathbf{w}_{t-1}, \tilde{\mathbf{y}}_{t-\tau:t-1}) p(\mathbf{w}_{t-1}|\mathbf{y}_{1:t-2}, \tilde{\mathbf{y}}_{1:t-2}) \quad (7)$$

$$p(y_t|\tilde{\mathbf{y}}_{1:t}, \mathbf{y}_{1:t-1}) = \int p(y_t|\mathbf{w}_t, \tilde{\mathbf{y}}_t) p(\mathbf{w}_t|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) d\mathbf{w}_t \quad (8)$$

$$p(\mathbf{w}_t|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) = \int p(\mathbf{w}_t|\mathbf{w}_{t-1}, \tilde{\mathbf{y}}_{t-\tau:t-1}) p(\mathbf{w}_{t-1}|\mathbf{y}_{1:t-2}, \tilde{\mathbf{y}}_{1:t-2}) d\mathbf{w}_{t-1} \quad (9)$$

$$(10)$$

And this recursions can be approximated by Monte Carlo simulation as described in the following section.

## 2.1 Non-linear filtering

The conditional (unknown) future density  $p(\mathbf{y}_{t+1}|\mathbf{y}_t)$  is approximated by applying Sequential Monte Carlo to the Monte Carlo empirical densities associated to the different predictive models. The algorithm is briefly described in the following.

- First, draw  $j$  independent values  $\mathbf{y}_{1:t+1}^j$ , with  $j = 1, \dots, M$  from  $p(\tilde{\mathbf{y}}_{s+1}|\mathbf{y}_{1:s})$ , with  $s = 1, \dots, t$ .
- Conditionally on  $\tilde{\mathbf{y}}_{1:t+1}^j$  obtain the particle sets  $\Xi_{1:t+1}^{i,j} = \{\mathbf{z}_{1:t+1}^{i,j}, \omega_t^{i,j}\}_{i=1}^N$ , with  $j = 1, \dots, M$ .



- Simulate  $\mathbf{y}_{t+1}^{i,j}$  from  $p(\mathbf{y}_{t+1}|\mathbf{z}_{t+1}^{i,j}, \tilde{\mathbf{y}}_{t+1}^j)$  and obtain

$$p_{N,M}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^N \omega_t^{i,j} \delta_{\mathbf{y}_{t+1}^{i,j}}(\mathbf{y}_{t+1})$$

For further details see Billio et al. [2010].

### 3 Empirical Application: AEX stock index

We forecast the one-month ahead log returns of the Amsterdam Exchange index (AEX) using two different models. The first alternative is a White Noise model (WN). This model assumes and thus forecasts that log returns are normally distributed with mean and standard deviation equal to the unconditional (up to time  $t$  for forecasting at time  $t + 1$ ) mean and standard deviation. WN is a standard benchmark to forecast stock returns since it implies a random walk assumption for prices, which is difficult to beat (see for example Welch and Goyal [2008]). The second alternative is a Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model:

$$y_t = \mu + \epsilon_t, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1), \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \epsilon_t^2 + \sum_{j=1}^p \sigma_{t-j}^2. \quad (11)$$

GARCH models are often employed in modeling financial time series that exhibit volatility clustering, i.e. periods of swings followed by periods of relative calm. We fix  $q = p = 1$  and estimate the model using Bayesian inference with an algorithm similar to Chen et al. [2005]. Finally, we apply our combination scheme (1)–(4) with time-varying weights (TVW), logistic-Gaussian dynamics and learning.

We evaluate the statistical accuracy of point forecasts given by the two individual models and the combination scheme in terms of the root mean square prediction error (RMSPE), and in terms of the correctly predicted percentage of sign (Sign Ratio) for the log percent stock index returns. We also evaluate the statistical accuracy of

the density forecasts in terms of the Kullback Leibler Information Criterion (KLIC). Specifically, the KLIC distance between the true density  $p(y_{t+1}|y_{1:t})$  of a random variable  $y_{t+1}$  and some candidate density  $p(\tilde{y}_{k,t+1}|y_{1:t})$  obtained from model  $k$  is defined as follows

$$\begin{aligned} \text{KLIC}_{k,t+1} &= \int p(y_{t+1}|y_{1:t}) \ln \frac{p(y_{t+1}|y_{1:t})}{p(\tilde{y}_{k,t+1}|y_{1:t})} dy_{t+1}, \\ &= \mathbb{E}_t[\ln p(y_{t+1}|y_{1:t}) - \ln p(\tilde{y}_{k,t+1}|y_{1:t})]. \end{aligned} \quad (12)$$

where  $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$  is the conditional expectation given the information set  $\mathcal{F}_t$  at time  $t$ . An estimate can be obtained from the average of the sample information,  $y_{t+1}, \dots, y_{\bar{t}+1}$ , on  $p(y_{t+1}|y_{1:t})$  and  $p(\tilde{y}_{k,t+1}|y_{1:t})$ :

$$\overline{\text{KLIC}}_k = \frac{1}{\bar{t}^*} \sum_{t=\underline{t}}^{\bar{t}} [\ln p(y_{t+1}|y_{1:t}) - \ln p(\tilde{y}_{k,t+1}|y_{1:t})]. \quad (13)$$

Even though we do not know the true density, we can still compare multiple densities,  $p(\tilde{y}_{k,t+1}|y_{1:t})$ . For the comparison of two competing models, it is sufficient to consider the Logarithmic Score (LS), which corresponds to the latter term in the above sum,

$$LS_k = -\frac{1}{\bar{t}^*} \sum_{t=\underline{t}}^{\bar{t}} \ln p(\tilde{y}_{k,t+1}|y_{1:t}), \quad (14)$$

for all  $k$  and to choose the model for which the expression in (14) is minimal, or as we report in our tables, the opposite of the expression in (14) is maximal.

Moreover, being an investor more interested in the economic value of a forecasting model than its precision, we test our conclusions in an active short-term investment exercise, with an investment horizon of one month. The investor's portfolio consists of the stock index and risk free bonds only.<sup>1</sup>

---

<sup>1</sup>The risk free asset is approximated by monthly averages of Money market rates reported by German banks. We collect them from the Bundesbank website.

At the end of each period  $t$ , the investor decides upon the fraction  $\alpha_{t+1}$  of her portfolio to be held in stocks for the period  $t + 1$ , based upon a forecast of the stock index return. We do not allow for short-sales or leveraging, constraining  $\alpha_{t+1}$  to be in the  $[0, 1]$  interval (see Barberis [2000]). The investor is assumed to maximize a power utility function with coefficient  $\gamma$  of relative risk aversion:

$$u(R_{t+1}) = \frac{R_{t+1}^{1-\gamma}}{1-\gamma}, \quad \gamma > 1, \quad (15)$$

where  $R_{t+1}$  is the wealth at time  $t + 1$ , which is equal to

$$R_{t+1} = R_t ((1 - \alpha_{t+1}) \exp(y_{f,t+1}) + \alpha_{t+1} \exp(y_{f,t+1} + \tilde{y}_{t+1})), \quad (16)$$

where  $R_t$  denotes initial wealth,  $y_{f,t+1}$  the 1-step ahead risk free rate and  $\tilde{y}_{t+1}$  the 1-step ahead forecast of the stock index return in excess of the risk free made at time  $t$ .

Without loss of generality we set initial wealth equal to one, i.e.  $R_0 = 1$ , such that the investor's optimization problem is given by

$$\max_{\alpha_{t+1} \in [0,1]} \mathbb{E}_t \left( \frac{((1 - \alpha_{t+1}) \exp(y_{f,t+1}) + \alpha_{t+1} \exp(y_{f,t+1} + \tilde{y}_{t+1}))^{1-\gamma}}{1-\gamma} \right),$$

How this expectation is computed depends on how the predictive density for the excess returns is computed. Following notation in section 2, this density is denoted as  $p(\tilde{y}_{t+1}|y_{1:t})$ . The investor solves the following problem:

$$\max_{\alpha_{t+1} \in [0,1]} \int u(R_{t+1}) p(\tilde{y}_{t+1}|y_{1:t}) d\tilde{y}_{t+1}. \quad (17)$$

We approximate the integral in (17) by generating with the SMC procedure  $MN$  equally weighted independent draws  $\{y_{t+1}^g, w_{t+1}^g\}_{g=1}^{MN}$  from the predictive density

$p(\tilde{y}_{t+1}|y_{1:t})$ , and then use a numerical optimization method to find:

$$\max_{\alpha_{t+1} \in [0,1]} \frac{1}{MN} \sum_{g=1}^{MN} \left( \frac{((1 - \alpha_{t+1}) \exp(y_{f,t+1}) + \alpha_{t+1} \exp(y_{f,t+1} + \tilde{y}_{t+1}^g))^{1-\gamma}}{1 - \gamma} \right) \quad (18)$$

We consider an investor who can choose between different forecast densities of the (excess) stock return  $y_{t+1}$  to solve the optimal allocation problem described above. We include three cases in the empirical analysis below and assume the investor uses alternatively the density from the WN model, the GARCH model or a density combination (DC) of the WN and SR densities. We apply here the DC scheme described in the previous section.

We evaluate the different investment strategies by computing the *ex post* annualized mean portfolio return, the annualized standard deviation, the annualized Sharpe ratio and the total utility. Utility levels are computed by substituting the realized return of the portfolios at time  $t + 1$  into (15). Total utility is then obtained as the sum of  $u(R_{t+1})$  across all  $t^* = (\bar{t} - \underline{t} + 1)$  investment periods  $t = \underline{t}, \dots, \bar{t}$ , where the first investment decision is made at the end of period  $\underline{t}$ . To compare alternative strategies we compute the multiplication factor of wealth that would equate their average utilities. For example, suppose we compare two strategies A and B. Wealths provided at time  $t + 1$  by the two resulting portfolios are denoted  $R_{A,t+1}$  and  $R_{B,t+1}$ , respectively. We then determine the value of  $\Delta$  such that

$$\sum_{t=\underline{t}}^{\bar{t}} u(R_{A,t+1}) = \sum_{t=\underline{t}}^{\bar{t}} u(R_{B,t+1} / \exp(r)). \quad (19)$$

Following Fleming et al. [2001], we interpret  $r$  as the maximum performance fee the investor would be willing to pay to switch from strategy A to strategy B. For comparison of multiple investment strategies, it is useful to note that – under a power utility specification – the performance fee an investor is willing to pay to switch from strategy A to strategy B can also be computed as the difference between

the performance fees of these strategies with respect to a third strategy C.<sup>2</sup> We use this property to infer the added value of strategies based on individual models and combination schemes by computing  $r$  with respect to three static benchmark strategies: holding stocks only ( $r_s$ ), holding a portfolio consisting of 50% stocks and 50% bonds ( $r_m$ ), and holding bonds only ( $r_b$ ).

Finally, the portfolio weights in the active investment strategies change every month, and the portfolio must be rebalanced accordingly. Hence, transaction costs play a non-trivial role and should be taken into account when evaluating the relative performance of different strategies. Rebalancing the portfolio at the start of month  $t + 1$  means that the weight invested in stocks is changed from  $\alpha_t$  to  $\alpha_{t+1}$ . We assume that transaction costs amount to a fixed percentage  $c$  on each traded dollar. Setting the initial wealth  $R_t$  equal to 1 for simplicity, transaction costs at time  $t + 1$  are equal to

$$c_{t+1} = 2c|\alpha_{t+1} - \alpha_t| \quad (20)$$

where the multiplication by 2 follows from the fact that the investor rebalances her investments in both stocks and bonds. The net excess portfolio return is then given by  $y_{t+1} - c_{t+1}$ . We apply a scenario with transaction costs of  $c = 0.1\%$  or 10 basis points.

Panel A in Table 3 reports static accuracy forecasting results. All the three approaches perform very similar in term of RMSPE and Sign Ratio, with the WN giving marginally smaller RMSPE and DC giving marginally higher Sign Ratio. DC seems, on the contrary, superior in terms of density forecasting: it gives the highest log score. Furthermore, Figure 1 plots the three density forecasts: the density forecasts of the two individual models are very similar and too wide, in particular the one produced by the WN. DC gives a smaller and more accurate interval.

---

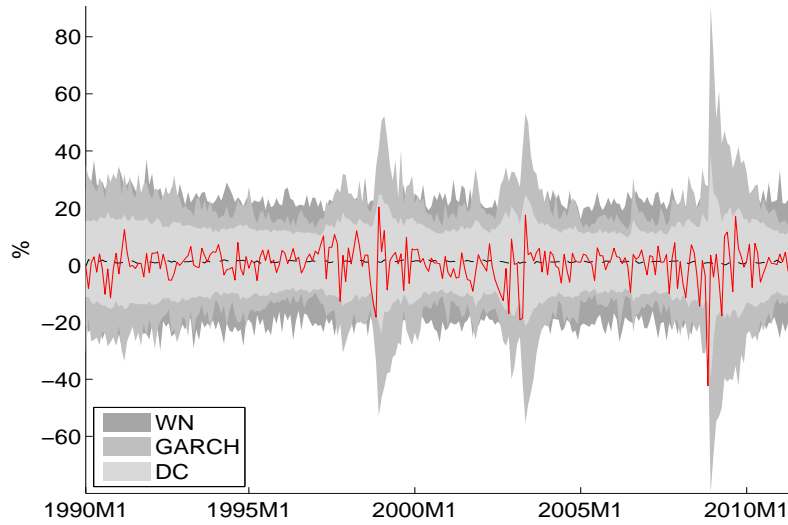
<sup>2</sup>This follows from the fact that combining (19) for the comparisons of strategies A and B with C,  $\sum_t u(R_{C,t+1}) = \sum_t u(R_{A,t+1}/\exp(r_A))$  and  $\sum_t u(R_{C,t+1}) = \sum_t u(R_{B,t+1}/\exp(r_B))$ , gives  $\sum_t u(R_{A,t+1}/\exp(r_A)) = \sum_t u(R_{B,t+1}/\exp(r_B))$ . Using the power utility specification in (15), this can be rewritten as  $\sum_t u(R_{A,t+1}) = \sum_t u(R_{B,t+1}/\exp(r_B - r_A))$ .

Table 1: Table: AEX index

	WN	GARCH	DC	WN	GARCH	DC
Panel A: Statical accuracy						
RMSPE	<b>6.442</b>	6.459	6.464			
SIGN	0.612	0.612	<b>0.616</b>			
LS	-3.391	-3.314	<b>-3.301</b>			
Panel B: no transaction costs						
		$\gamma=4$			$\gamma=6$	
Mean	0.695	<b>0.700</b>	0.699	0.694	0.697	<b>0.698</b>
St dev	0.415	<b>0.415</b>	0.416	0.336	0.337	<b>0.336</b>
SR	0.002	<b>0.015</b>	0.013	0.001	0.008	<b>0.011</b>
Utility	-84.10	<b>-84.10</b>	-84.10	-50.12	-50.11	<b>-50.11</b>
r_s	132.9	<b>133.3</b>	133.3	251.7	251.8	<b>252.0</b>
r_m	17.92	<b>18.35</b>	18.34	39.30	39.48	<b>39.62</b>
r_b	1.338	<b>1.774</b>	1.762	0.812	0.992	<b>1.135</b>
Panel C: transaction costs						
		$\gamma=4$			$\gamma=6$	
Mean	0.695	<b>0.700</b>	0.699	0.694	0.697	<b>0.698</b>
St dev	0.415	0.415	<b>0.416</b>	0.336	0.337	<b>0.336</b>
SR	0.002	<b>0.014</b>	0.012	0.001	0.008	<b>0.011</b>
Utility	<b>-84.13</b>	-84.14	-84.15	<b>-50.13</b>	-50.14	-50.15
r_s	<b>131.7</b>	131.0	130.2	<b>250.9</b>	250.4	250.1
r_m	<b>16.38</b>	15.64	14.85	<b>38.19</b>	37.62	37.30
r_b	<b>0.186</b>	-0.553	-1.341	<b>0.092</b>	-0.480	-0.793

*Note:* In Panel A the root mean square prediction error (RMSPE), the correctly predicted sign ratio (SIGN) and the Logarithmic Score (LS) for individual models and combination schemes in forecasting the one month ahead AEX index return over the period January 1990 - March 2011. WN, GARCH and DC denote strategies based on excess return forecasts from the White Noise model, the Garch model and our density combination scheme in equations (1)–(4). In Panel B the annualized percentage point average portfolio return and standard deviation, the annualized Sharpe ratio (SR), the final value of the utility function, and the annualized return in basis points that an investor is willing to give up to switch from the passive stock (s), mixed (m), or bond (b) strategy to the active strategies and short selling and leveraging restrictions are given. In Panel C the same statistics as in Panel B are reported when transaction costs  $c = 10$  basis points are assumed. The results are reported for two different risk aversion coefficients  $\gamma = (4, 6)$ .

Figure 1: Prediction densities for AEX

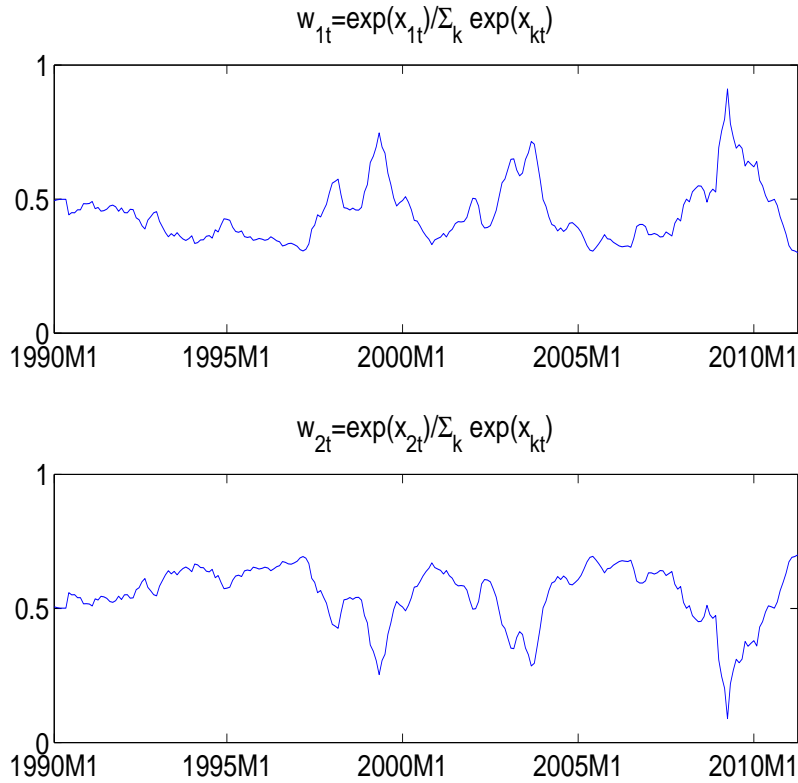


*Note:* The figure presents the (99%) interval forecasts given by the White Noise benchmark model (WN), the Garch model and our density combination scheme (DC). The red solid line shows the realized values for AEX percent log returns, for each out-of-sample observation.

Figure 2 shows the combination weights with learning. The median weights vary over time depending on the volatility level of the predicted variable. The GARCH model has larger weight (between 0.5 and 0.7 in the figure) in periods of slowly changing or constant volatility. The weight, however, reduces dramatically (down to 0.1) in periods of exploding volatility such as after the Russian crisis in October 1998, after the large turmoil in the second semester of 2002, and after the failure of Lehman Brothers in September 2008.

The results for the asset allocation exercise strengthen previous statistical accuracy evidence. Panel B in Table 3 reports results for two different risk aversion coefficients,  $\gamma = (4, 6)$ . The performance fees are all positive for both investors and all approaches, with the GARCH model and the DC giving the highest result for coefficient aversions of 4 and 6 respectively. Difference are larger in term of Sharpe ratio. Adding transaction costs changes results: the WN is the only one to provide positive fees and investing all the money in the bond asset seems a passive strategy difficult to beat. Considering the recent crises with a dramatic fall in stock returns,

Figure 2: Combination weights for AEX forecasts



*Note:* Time-varying weights for WN (top panel) and Garch (bottom panel) models for AEX predictions.

our findings are not completely surprising. Results are also different than Billio et al. [2010] where DC gives high performance. The high correlations between our individual forecasts and their poor performance in various metrics can explain such discrepancy.

## 4 Conclusion

This paper combines different predictive densities to forecast financial time series using a method based on a distributional state-space representation of the prediction model and of the combination scheme and on a Bayesian filtering of the optimal weights proposed by Billio et al. [2010]. An application to forecast financial stock



index shows that our methodology improves statistical accuracy, in particular in terms of density forecasting. This does not guarantee higher investment performance.

The results indicate, however, that our proposed approach has wider application than the illustrative example used in the present paper. We intend to pursue our line of research further in future work.

## References

- N. Barberis. Investing for the Long Run When Returns are Predictable. *Journal of Finance*, 55:225–264, 2000.
- J. M. Bates and C. W. J. Granger. Combination of Forecasts. *Operational Research Quarterly*, 20:451–468, 1969.
- M. Billio, R. Casarin, F. Ravazzolo, and H.K. Van Dijk. Combining predictive densities using bayesian filtering with applications to us economics data. Norges Bank Working Paper No. 2010/29, 2010.
- C.W.S. Chen, M.K.P. So, and MT Chen. A bayesian threshold nonlinearity test for financial time series. *Journal of Forecasting*, 24(1):61–75, 2005.
- F. X. Diebold and P. Pauly. Structural change and the combination of forecasts. *Journal of Forecasting*, 6:21–40, 1987.
- J. Fleming, C. Kirby, and B. Ostdiek. The Economic Value of Volatility Timing. *Journal of Finance*, 56:329–352, 2001.
- C. W. J. Granger. Invited review combining forecasts - twenty years later. *Journal of Forecasting*, 8:167–173, 2006.
- C. W. J. Granger and R. Ramanathan. Improved Methods of Combining Forecasts. *Journal of Forecasting*, 3:197–204, 1984.

- J. Groen, R. Paap, and F. Ravazzolo. Real-Time Inflation Forecasting in a Changing World. Technical Report, 2009.
- J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14:382–417, 1999.
- L. Hoogerheide, R. Kleijn, R. Ravazzolo, H. K. van Dijk, and M. Verbeek. Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights. *Journal of Forecasting*, 29(1-2):251–269, 2010.
- N. Terui and H. K. van Dijk. Predictability in the Shape of the Term Structure of Interest Rates. *International Journal of Forecasting*, 18:421–438, 2002.
- I. Welch and A. Goyal. A Comprehensive Look at the Empirical Performance of Equity Premium prediction. *Review of Financial Studies*, 21(4):253–303, 2008.