DISCUSSION PAPER SERIES

No. 9096

OPTIMAL COMBINATION OF SURVEY FORECASTS

Cristina Conflitti, Christine De Mol and Domenico Giannone

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9096.asp**

ISSN 0265-8003

OPTIMAL COMBINATION OF SURVEY FORECASTS

Cristina Conflitti, ECARES
Christine De Mol, ECARES
Domenico Giannone, ECARES and CEPR

Discussion Paper No. 9096
August 2012

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9096

August 2012

ABSTRACT

Optimal Combination of Survey Forecasts*

We consider the problem of optimally combining individual forecasts of gross domestic product (GDP) and inflation from the

Survey of Professional Forecasters (SPF) dataset for the Euro Area. Contrary to the common practice of using equal combination weights, we compute optimal weights which minimize the mean square forecast error (MSFE) in the case of point forecasts and maximize a logarithmic score in the case of density forecasts. We show that this is a viable strategy even when the number of forecasts to combine gets large, provided we constrain these weights to be positive and to sum to one. Indeed, this enforces a form of shrinkage on the weights which ensures good out-of-sample performance of the combined forecasts.

JEL Classification: C22 and C53
Keywords: forecast combination, forecast evaluation, high-dimensional data, real-time data, shrinkage and survey of professional forecasters

Cristina Conflitti
PhD student- ECARES
Universite Libre de Bruxelles
50, avenue Roosevelt
CP 114, B1050 Bruxelles
BELGIUM

Email: cristina.conflitti@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=171160

Christine De Mol
ECARES and Department of Mathemathics
University Libre de Bruxelles
ULB Campus Plaine CP 217
Boulevard du Triomphe
B-1050 Brussels
BELGIUM

Email: demol@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=161494

Domenico Giannone
Université Libre de Bruxelles
ECARES
50 Av. F.D. Roosevelt, CP 114
B-1050 Bruxelles
BELGIUM

Email: dgiannon@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=154334

*The views expressed in the article do not involve the responsibility of the Bank of Italy.

Submitted 11 August 2012

1 Introduction

The idea of linearly combining individual forecasts provided by different sources in order to improve accuracy and reliability is quite an old one. There is a vast literature on the subject, advocating the usefulness of forecast combination methods both from a theoretical point of view and on the basis of the results of empirical studies. The forecasts to combine can be either judgemental, provided e.g. by individual forecasters participating to surveys, or else provided by different quantitative models. In the present paper we focus on survey data, and in particular on the ECB Survey of Professional Forecasters (SPF) dataset. This survey has been conducted by the European Central Bank (ECB) at a quarterly frequency since the inception of the European Monetary Union (EMU). Survey participants are experts affiliated with financial and non-financial European institutions. They are asked to provide point and density forecasts for GDP, inflation and unemployment at different horizons. A detailed description of the survey is contained in the papers by Garcia (2003) and Bowles, Friz, Genre, Kenny, Meyler and Rautanen (2007, 2010).

A simple and widely used combination method consists in simply averaging all available forecasts of a given variable, attributing equal weights to the individual predictions. However, the idea of determining optimal combination weights that minimize some objective criterion or cost function seems more appealing. When combining point forecasts, a natural target to minimize is the mean square forecast error (MSFE), i.e. the variance of the combination around the variable to be predicted. In practice, this minimization can be performed over some available historical periods, so that the optimal weights minimize an empirical least-squares criterion. In economics, this idea dates back to Bates and Granger (1969) and Granger and Ramanathan (1984) and has been the subject of a great variety of developments including the use of different optimality criteria, of time-varying weights, of nonlinear combination schemes, etc. For a review of the literature, we refer to the survey papers by Clemen (1989) and by Timmermann (2006). More recently, a similar approach has been advocated for density forecasts using combination weights minimizing the so-called logarithmic score (Hall and Mitchell, 2007, Geweke and Amisano, 2011).

A closer look at this literature shows that, when dealing with applications, only a rather small number of individual forecasts are considered for optimal combination whereas optimality seems to be given up as soon as this number becomes large. For example, the recent papers by Sloughter, Gneiting and Raftery (2010) and by Geweke and Amisano (2011, 2012) deal with combinations of just a handful of individual forecasts. In other papers considering larger combinations, the weights are taken as equal or are assigned on the sole basis of the previous performance of each forecaster, ignoring their mutual dependence and hence eliminating the need for optimization (two important recent examples are the papers by Clark and McCracken (2010) for point forecasts and by Jore, Mitchell and Vahey (2010) for densities). This strategy appears to be justified empirically by the fact that the resulting simple averaging schemes tend to outperform more sophisticated ones. Such phenomenon is usually referred to as the "forecasting combination puzzle" and has been recently documented for our dataset by Genre, Kenny, Meyer and Timmermann (2010), who show that the simple equal-weight averages constitute a benchmark which is very hard to improve upon, at least for GDP growth and unemployment rate. This explains why this practice still prevails today among institutions such as ECB. Interestingly, a similar phenomenon has been observed in portfolio optimization by DeMiguel, Garlappi and Uppal (2009), a problem which shares with forecast combination the idea due to Markowitz of exploiting diversification to decrease risk/variance.

In the present paper we show that there is no need to give up optimality when going to a high-dimensional setting, i.e. when combining a large number of forecasts. The reason why previous works either stick to small combinations or rely for large datasets to simplified covariance modelling is most likely related to two fundamental difficulties: (i) the presence of finite-sample errors and numerical instabilities in the estimation of the weights (see e.g. Smith and Wallis, 2009); (ii) the need for solving the resulting high-dimensional optimization problem in an efficient computational way. Both for the cases of point and density forecast combinations, we argue that the determination of the optimal weights is stabilized by the constraint that they should be positive and add up to one and we show that the computation of these optimal weights is easily implementable using iterative algorithms that can handle efficiently a large number of forecasts.

In Section 2, we deal with the combination of point forecasts, defining the optimal weights as minimizers of the MSFE over some historical period, imposing the constraints that these weights are positive (more precisely nonnegative) and sum to one. Hence the optimal combination problem reduces to a (possibly high-dimensional) constrained least-squares regression problem where the complete covariance structure between forecasters is taken into account. We show that the combined use of these two – rather natural – constraints on the weights is essential for a proper formulation of the problem, in the sense that it enforces an implicit shrinkage of the weights which makes their computation stable with respect to errors in the data. In fact the problem turns out to be analogous to the determination of no-short Markowitz portfolios, i.e. portfolios for which the weights are constrained to be nonnegative. As established by Brodie, Daubechies, De Mol, Giannone and Loris (2009), such portfolios are a special case of a larger family of sparse and stable portfolios derived through a constrained "lasso" regression problem. We recall that in the so-called "lasso" regression (Tibshirani, 1996), the least-squares objective function is modified by adding a penalty term proportional to the sum of the absolute values of the weights (i.e. to the L1-norm of the weight vector). This analogy allows to borrow from the work by Brodie et al. (2009) the efficient algorithm proposed in that paper to compute the optimal weights. Moreover, these considerations imply that the weight vector solving our optimization problem is sparse, i.e. that many of the weights are exactly zero. This fact naturally provides a selection of ideally diversified forecasters to be optimally combined in order to minimize the MSFE. Moreover, due to the constraints fixing its L1-norm to be one and to the resulting nonlinear "shrinkage" provided by the lasso technique, the optimal weight vector is expected to be stable with respect to small fluctuations of the data even for large panels of forecasters, which is not generally the case for ordinary least-squares estimates in high-dimensional situations (for more details about this point, we refer to the paper by Brodie et al. (2009)).

The idea of optimal combination can be extended to the case of density forecasts, but the similarity between two densities is usually measured by means of the so-called Kullback-Leibler divergence or Kullback-Leibler Information Criterion (KLIC), instead of the least-squares distance. Besides, in the case of survey data, we miss a target density since only the realized value of the variable to forecast, say, gross domestic product (GDP) or inflation, is available. Then, as proposed by Hall and Mitchell (2007), we show in Section 3 that the optimal weights can be obtained by maximizing a logarithmic score function, under the constraints that they are nonnegative and sum to one, which ensures that the combination of densities is still a proper density. To compute such weights we derive a simple iterative algorithm which scales well with the dimension of the panel, i.e. allows to handle large datasets.

Section 4 contains our empirical analysis. The SPF point and density forecasts for GDP growth and inflation are optimally combined as described above. They are compared with the equal-weight combinations used by the ECB to summarize the results of each round of the survey. The evaluation is performed by means of a real-time out-of-sample forecasting exercise.

Finally, Section 5 contains the conclusions of our work as well as some pointers to other potential applications of our combination framework.

2 Optimal combination of point forecasts

As in the paper by Granger and Ramanathan (1984), we address the problem of determining the optimal combination weights for point forecasts as a least-squares regression problem (hence using the full covariance structure between forecasters) but, in addition, we impose the two constraints that the weights should be nonnegative and sum to one, constraints which make the ordinary least squares (OLS) estimation feasible and stable, as shown below.

We denote by y_{t+h} the variable to be forecast at time t, assuming that the desired forecast horizon is h, and we suppose to have at our disposal N forecasters, each providing at time t a forecast $\hat{y}_{i,t+h|t}$, based on the information available at time t. We form with these individual forecasts $\hat{y}_{i,t+h|t}$, $i = 1, \ldots, N$ the $N \times 1$-dimensional vector $\hat{\mathbf{y}}_{t+h|t}$. We want to linearly combine these N forecasts by means of time-independent weights w_i, $i = 1, \ldots, N$ with $\sum_{i=1}^{N} w_i = 1$ and $w_i \geq 0$, which we put into the $N \times 1$ vector $\mathbf{w}$. The goal is to reduce the forecast error $e_{t+h|t}(\mathbf{w}) = y_{t+h} - \mathbf{w}'\hat{\mathbf{y}}_{t+h|t}$ achieved by the combination $\mathbf{w}'\hat{\mathbf{y}}_{t+h|t} \equiv \sum_{i=1}^{N} w_i \hat{y}_{i,t+h|t}$ and to minimize it according to some criterion, which we choose here to be the mean square forecast error (MSFE). Accordingly, the optimal weight vector $\mathbf{w}_{\mathrm{OPT}}$ is defined as the vector such that

$$\mathbf{w}_{\mathrm{OPT}} = \operatorname*{argmin}_{\mathbf{w}} \mathrm{E}[(y_{t+h} - \mathbf{w}'\hat{\mathbf{y}}_{t+h|t})^2] \equiv \mathrm{E}[e^2_{t+h|t}(\mathbf{w})] \tag{1}$$

$$\text{subject to} \quad \sum_{i=1}^{N} w_i = 1 \text{ and } w_i \geq 0,$$

.

In empirical applications, the expectation has to be replaced by a sample mean over some historical period for which both the forecasts and the realization of the real variable are available. Hence the minimization problem becomes:

$$\hat{\mathbf{w}}_{\mathrm{OPT}} = \operatorname*{argmin}_{\mathbf{w}} \sum_{t=1}^{T-h} (y_{t+h} - \mathbf{w}'\hat{\mathbf{y}}_{t+h|t})^2 \tag{2}$$

$$\text{subject to} \quad \sum_{i=1}^{N} w_i = 1 \text{ and } w_i \geq 0,$$

assuming that the variable y_t is observed for $t = 1, \ldots, T$. This optimal vector of weights $\hat{\mathbf{w}}_{\mathrm{OPT}}$ can then be computed and used to form the combined forecast $(\hat{\mathbf{w}}_{\mathrm{OPT}})'\hat{\mathbf{y}}_{T+h|T}$ of the variable y_t at time $t = T + h$.

This problem bears strong similarity with the problem of determining minimum-variance (i.e. without target-return constraint) no-short (i.e. with nonnegative weights) portfolios, the vector of forecasts being replaced with a vector of returns. Besides, let us remark that the above minimization problem is equivalent to the following one

$$\hat{\mathbf{w}}_{\mathrm{OPT}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left[\sum_{t=1}^{T-h} (y_{t+h} - \mathbf{w}'\hat{\mathbf{y}}_{t+h|t})^2 + \lambda \sum_{i=1}^{N} |w_i| \right] \tag{3}$$

$$\text{subject to } \sum_{i=1}^{N} w_i = 1 \text{ and } w_i \geq 0,$$

where λ is a positive parameter. Indeed, under the two constraints on the weights, the added term, which is simply the unit L1-norm of the weight vector multiplied by λ, is just a constant. Hence the problem amounts to a so-called "lasso" regression (Tibshirani, 1996) with two additional constraints. The resulting optimization problem is convex and, contrary to OLS, well defined even when N is larger than T. This lasso strategy was used in the paper by Brodie et al. (2009) to construct a family of Markowitz-like portfolios which are sparse, i.e. present few active positions (corresponding to the non-zero weights) and also stable, i.e. with weight values not oversensitive to errors in the data. Moreover, in that paper, it was shown that the no-short portfolios – being special cases of the family obtained when omitting the positivity constraint and appropriately tuning the parameter λ – are sparse, i.e. that many weights are exactly zero, which in our case means that only a small number of forecasts will be selected (active) to form the combined forecast. Let us remark that the number of selected forecasters is here entirely determined by the data and cannot be tuned by means of the value of λ as it is the case for the number of active assets of the sparse portfolios when short positions (negative weights) are allowed (for a theoretical discussion of this point, see Brodie et al., 2009). The sparsity property of the solution of our optimization problem can be viewed either as an advantage or as a drawback. This point will be discussed further in Section 4, in connection with our empirical results.

From a computational point of view, the optimization problem (2) is a quadratic program which could be solved by any appropriate algorithm. However, not all such solvers scale well with the dimension N nor do allow to determine the number of active forecasters since they do not distinguish between true zero or approximately zero values for the weights. Therefore we have used the algorithm developed by Brodie et al. (2009) to compute the sparse Markowitz portfolios. It is based on the LARS algorithm proposed by Efron, Hastie, Johnstone and Tibshirani (2004) for lasso regression, but modified in order to enforce linear equality constraints. For a detailed description of this constrained LARS algorithm, we refer the reader to the paper by Brodie et al. (2009).

3 Optimal combination of density forecasts

A density forecast is an estimate of the future probability distribution of a variable of interest. Compared to point forecasts, the literature on density forecast combination is less abundant but bears testimony of the practical relevance of the problem; see e.g. the reviews by Genest and Zidek (1986), Tay and Wallis (2000) and Wallis (2011), as well as the papers by Diebold, Tay and Wallis (1999) and Clements and Harvey (2007).

A natural and widely used measure of similarity between probability densities is the Kullback-Leibler Information Criterion (KLIC) adopted by Hall and Mitchell (2007). Accordingly, the problem of optimal combination of density forecasts can be formulated as follows. Let us denote by $\hat{\mathbf{p}}(y_{t+h})$ the vector of the N density forecasts $\hat{p}_i(y_{t+h})$ of a given variable y_{t+h} made by the individual forecasters $(i = 1, \ldots, N)$ at time t over some horizon h, and write the combined density as:

$$\hat{p}(y_{t+h}) = \mathbf{w}'\hat{\mathbf{p}}(y_{t+h}) \equiv \sum_{i=1}^{N} w_i \hat{p}_i(y_{t+h}), \tag{4}$$

assuming, as in Section 2, that the weights are nonnegative and sum to one, which ensures that the combined density is still a probability density.
The Kullback-Leibler divergence or Information Criterion between the true density $p(y_{t+h})$ and the combined density $\hat{p}(y_{t+h})$ is defined as:

$$KLIC = \int p(y_{t+h}) \ln \frac{p(y_{t+h})}{\hat{p}(y_{t+h})} dy_{t+h} = E[\ln p(y_{t+h}) - \ln \hat{p}(y_{t+h})]. \tag{5}$$

A consistent estimate of (5) is given by its sample average. In the case of survey forecasts, we miss the reference target density $p(y_{t+h})$, but averaging its logarithm over the sample yields a constant independent of the combination weights. Hence the problem of determining the optimal weights from empirical data reduces to the maximization of the (concave) cost function:

$$\Phi(\mathbf{w}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \ln \hat{p}(y_{t+h}) \tag{6}$$

i.e. of the average logarithmic score of the combined density (4) over the available sample $t = 1, \ldots, T-h$ (as proposed by Hall and Mitchell, 2007).

To compute the corresponding optimal combination weights, we devised the following simple iterative algorithm. Let us define the $(T-h) \times N$ matrix $\hat{P}$ with nonnegative elements $\hat{P}_{ti} = \hat{p}_i(y_{t+h})$. Then (6) can be rewritten as $\Phi(\mathbf{w}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \ln(\hat{P}\mathbf{w})_t$. We denote by $\hat{\mathbf{w}}_{\mathrm{OPT}}$ the maximizer of $\Phi(\mathbf{w})$ subject to the constraints, $w_i \geq 0$ and $\sum_{i=1}^{N} w_i = 1$. To take into account the constraint that the weights should sum to one, let us introduce a Lagrange multiplier λ and maximize the following function:

$$\Phi_\lambda(\mathbf{w}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \ln(\hat{P}\mathbf{w})_t - \lambda \sum_{i=1}^{N} w_i. \tag{7}$$

To perform the maximization problem, we introduce the following "surrogate" cost function depending on a vector $\mathbf{a}$ of arbitrary weights:

$$\Psi_\lambda(\mathbf{w};\mathbf{a}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \sum_{i=1}^{N} \frac{\hat{P}_{ti} a_i}{\sum_{l=1}^{N} \hat{P}_{tl} a_l} \ln \left(\frac{w_i}{a_i} \sum_{l=1}^{N} \hat{P}_{tl} a_l \right) - \lambda \sum_{i=1}^{N} w_i \ . \tag{8}$$

This surrogate has the following properties:

- (i) $\Psi_\lambda(\mathbf{a};\mathbf{a}) = \Phi_\lambda(\mathbf{a})$ for any $\mathbf{a}$,
- (ii) $\Psi_\lambda(\mathbf{w};\mathbf{a}) \leq \Phi_\lambda(\mathbf{w})$ for any $\mathbf{a}$ and any $\mathbf{w}$.

The first assertion is straightforward whereas the second follows immediately from the inequality expressing the concavity of the logarithmic function

$$\sum_{i=1}^{N} b_{ti} \ln(x_{ti}) \leq \ln \left(\sum_{i=1}^{N} b_{ti} x_{ti} \right) \tag{9}$$

with $b_{ti} = \frac{\hat{P}_{ti} a_i}{\sum_{l=1}^{N} \hat{P}_{tl} a_l}$ and $x_{ti} = \frac{w_i}{a_i} \sum_{l=1}^{N} \hat{P}_{tl} a_l$, noticing that $\sum_{i=1}^{N} b_{ti} = 1$ for all t.

We can then define the iterative algorithm as

$$\mathbf{w}_\lambda^{(k+1)} = \mathrm{argmax}_{\mathrm{w}} \Psi_\lambda(\mathbf{w}; \mathbf{w}_\lambda^{(\mathrm{k})}) \tag{10}$$

which yields a monotonic increase of Φ_λ, i.e. $\Phi_\lambda(\mathbf{w}_\lambda^{(k+1)}) \geq \Phi_\lambda(\mathbf{w}_\lambda^{(k)})$, as follows from the properties (i) and (ii). Indeed, $\Phi_\lambda(\mathbf{w}_\lambda^{(k+1)}) \geq \Psi_\lambda(\mathbf{w}_\lambda^{(k+1)}; \mathbf{w}_\lambda^{(k)}) \geq \Psi_\lambda(\mathbf{w}_\lambda^{(k)}; \mathbf{w}_\lambda^{(k)}) = \Phi_\lambda(\mathbf{w}_\lambda^{(k)})$.

Notice that the maximization of (8) is easy since the surrogate is separable: it is the sum of N terms, each depending only on a single weight. Equating to zero the derivatives of $\Psi_\lambda(\mathbf{w}; \mathbf{w}_\lambda^{(k)})$ with respect to each weight, we find that its maximizer is given by $w_{\lambda,i} = \frac{1}{\lambda} \sum_{t=1}^{T} b_{ti}$. To determine λ, we use the constraint $1 = \sum_{i=1}^{N} w_{\lambda,i} = \frac{1}{\lambda} \sum_{t=1}^{T-h} \sum_{i=1}^{N} b_{ti} = \frac{T-h}{\lambda}$. Hence $\lambda = T - h$ and the iterative algorithm (10) becomes

$$w_i^{(k+1)} = w_i^{(k)} \frac{1}{T-h} \sum_{t=1}^{T-h} \frac{\hat{P}_{ti}}{\sum_{l=1}^{N} \hat{P}_{tl} w_l^{(k)}} \tag{11}$$

replacing a_i by $w_i^{(k)}$ in the expression of b_{ti}.

The nonnegativity constraint for the weights is automatically satisfied at each iteration provided that the algorithm is initialized with positive weights summing to one, e.g. with $w_i^{(0)} = 1/N$. Due to the monotonic increase of the cost function, the iterates $\mathbf{w}^{(k)}$ are expected to converge to the maximizer $\hat{\mathbf{w}}_{\mathrm{OPT}}$ of (6), subject to the two constraints $w_i \geq 0$ and $\sum_{i=1}^{N} w_i = 1$. However, a detailed analysis of the convergence properties of this algorithm is beyond the scope of the present paper.

The algorithm can be terminated by means of an appropriate stopping criterion, e.g. by stopping as soon as the components of two successive iterates do not differ by more that some predefined accuracy tolerance.

A major advantage of the previous algorithm is that it is very simple to implement and scales well with the cross-sectional dimension of the problem, and hence is not limited to the combination of a small number of density forecasts.

4 Empirical analysis

Since 1999 the European Central Bank (ECB) has conducted the Survey of Professional Forecasters (SPF) which is a quarterly survey of expectations for some of the Euro Area key macroeconomic variables: HICP (Harmonised Index of Consumer Prices) inflation, real GDP growth rate and unemployment rate.

For each variable and for different horizons, the forecasters are asked to report, together with point forecasts, a probability distribution or density forecast, by allocating probabilities to ranges of possible outcomes. Usually, the lower bottom interval and the upper interval of the range are open bins, and the interior bins have equal lengths of 0.5. The results of the survey are collected in the second half of the first month of each quarter.

The survey has been conducted between 1999Q1 and 2001Q3 in the middle month of each quarter, i.e. in February, May, August and November. Since 2001Q4 the survey has been shifted to the first month of the quarter, i.e. it has been and is still currently conducted in January, April, July and October. The questionnaire is sent to the panelists just after the Harmonized Index of Consumer Prices (HICP) release, that is in the third week of the month before the survey. Hence the forecasts are collected in the second half of the first month of each quarter. A detailed description of the survey is provided in the papers by Garcia (2003) and by Bowles et al. (2007, 2010). For background information, including the sample questionnaire, see also the ECB website.[1]

The dataset is characterized by several missing data due to the entering and exiting of forecasters in the panel. Since the implementation of the optimal combination scheme requires a full panel without missing observations, we pre-filter the data in order to create a balanced panel. First, we exclude those forecasters with more than 25 missing survey rounds. Then, for each of the 63 remaining individual forecasters, the unreported point forecasts are filled with the most recent forecast he provided. Missing observations in the first survey round are replaced with the average opinion of the respondents.[2] For unreported density forecasts, a uniform density is taken as replacement.[3]

We focus on expectations about the year-on-year GDP growth and year-on-year HICP inflation at an horizon of one year ahead of the latest available data for the respective variables. For example, in the first survey in the first quarter of 2008 (sent out after the official release of the December 2007 figure for HICP inflation and of the 2007Q3 figure for GDP), the questionnaire asked for the expected year-on-year inflation rate in December 2008 and the year-on-year GDP growth in 2008Q3. In the second survey, the 2008Q2 SPF (sent

[1] http://www.ecb.int/stats/prices/indic/forecast/html/index.en.html

[2] As an alternative to this "naive" method, we could have balanced the dataset assuming an autoregressive process of order one ($AR(1)$), as in Genre et al. (2010), or a principal component method (PC_M) as proposed in Stock and Watson (2002). We have investigated the sensitivity of the empirical results to the choice of the replacement technique and concluded that they were not significantly affected by the filling method.

[3] When a response is missing, to each bin we assign the value $1/h$, where h is the number of bins provided in the questionnaire.

out after the release of the March 2008 HICP figure and of the 2007Q4 figure for GDP), the questionnaire asked for the expected year-on-year inflation rate in March 2009, and the year-on-year GDP growth in 2008Q4.

We carry out an analysis using 48 survey rounds, from 1999Q1 until 2010Q4. For each quarter we estimate the weights using the most recent 5 years of data as they were available at the time the questionnaire was sent out by the ECB to the professional forecasters. Since data on GDP and inflation are subject to revisions, we use the Euro Area real-time database ("RTDB"; see Giannone, Henry, Lalik and Modugno, 2010) to match the survey data with the information that was available to the forecasters at the time they submitted their projections. Predictions are compared with the first official release published on the ECB Monthly Bulletin.[4]

Figure 1 compares the outturn (red solid line) with the predictions obtained by forecast combination. We report the 68% bands (shaded area) and the median (line with blue stars). We also plot, as a green dotted line, the results obtained for point forecasts with the corresponding combination scheme. The first column corresponds to the year-on-year GDP growth and second column to the year-on-year HICP inflation. The top panel refers to the forecasts computed using the optimal combination weights defined above, whereas the bottom panel refers to the naive combination reported by the European Central Bank (ECB), where equal weights are attributed to all respondents in a given round.

The accuracy of the point forecasts is evaluated by computing the mean square forecast error. We report in Table 1 the MSFE for the optimal scheme and, for the ECB scheme, the difference in MSFE with respect to the optimal one.

For the differences in MSFE we report heteroscedasticity-and-autocorrelation corrected (HAC) standard deviations. The implied t-statistics provides a valid test for equal predictive accuracy since the optimal weights are computed using a rolling scheme (see Giacomini and White, 2004).

Analyzing these results, we see that for inflation the optimal combination is significantly more accurate than the equal-weight forecast combination used by the ECB. For GDP, however, the optimal weights perform slightly worse

[4]The first questionnaire considered was sent out after the official release of the December 1998 figure for HICP inflation and of the 1998Q3 figure for GDP. The last questionnaire considered was sent out after the official release of the March 2010 figure for HICP inflation and of the 2009Q4 figure for GDP. Since we consider the forecasts one year ahead of the last available data, the evaluation sample runs from 1999Q3 until 2010Q4 for GDP and from December 1999 until December 2010 for inflation.

than the ECB weights but the difference in forecast accuracy is not statistically significant.

Finally, we observe that the optimal weight vector is quite sparse: the number of active forecasters for GDP varies from 1 to 6 and for inflation from 2 to 5, which is quite small compared to the total number of forecasters. As explained in Section 2, the two constraints on the weights enforce sparsity, i.e. the presence of many zero weights, which amounts to selecting few forecasters. Let us remark that the selected set of forecasters may nevertheless vary from step to step in the rolling scheme, especially in the presence of high data collinearity; such instability in the selection is a known drawback of lasso regression (see e.g. the paper by De Mol, Giannone and Reichlin (2008)). As for the number of assets selected in no-short portfolios, the number of active forecasters cannot be tuned. A way around this limitation would be to allow for negative weights (analogous to short positions in portfolios) and to use a lasso regression scheme with a tunable parameter λ able to regulate the sparsity of the weight vector, as done by Brodie et al. (2009).

Table 1: MSFE for point forecasts

	OPT	(ECB – OPT)
GDP	2.80	-0.04
		(0.03)
Inflation	0.77	0.07
		(0.01)

The first column displays the MSFE corresponding to the optimal combination. The second column displays the difference in MSFE between the ECB and optimal combinations. The standard deviations for these differences, corrected for autocorrelation and heteroscedasticity (HAC), are reported between brackets.

The accuracy of the density forecasts is evaluated by computing the average logarithmic score. Remember that the optimal weights maximize the logarithmic score, but to facilitate the comparison with the case of point forecast, we report in the first column of Table 2 the loss or "minus-log" score for the optimal combination, i.e. minus the logarithmic score, so that the lower the value the better the forecast performance.

In Table 2, we report in the first column the average minus-log score of the optimal weighting scheme (OPT) and in the second column the average and the HAC standard deviations of the differences between the minus-log scores for the ECB and OPT schemes. The implied t-statistics provides a Likelihood Ratio Test of equal predictive ability (see Amisano and Giacomini, 2007). We see from Table 2 that the average minus-log score obtained via the optimal scheme outperforms the ECB combination method for both GDP growth and HICP inflation but that the difference is significant only for inflation.

Table 2: Average minus-log score for density forecasts

	OPT	(ECB – OPT)
GDP	2.08	0.01
		(0.02)
Inflation	1.93	0.24
		(0.10)

The first column displays the average minus-log score for the optimal combination. The second column refers to the ECB scheme in terms of difference with respect to the optimal scheme. The standard deviations for these differences, corrected for autocorrelation and heteroscedasticity (HAC), are reported between brackets.

5 Conclusion

We have shown that optimal combinations of point and density forecasts is a viable strategy even in a high-dimensional setting. The optimal point forecasts, defined as to minimize the Mean Square Forecast Error, can be easily computed from empirical data by a constrained least-squares regression which turns out to be a special case of a lasso regression. For the density forecasts, the optimal weights, defined as to maximize a logarithmic score deriving from the Kullback-Leibler Information Criterion, can also be easily computed from empirical data by means of a simple iterative algorithm.

When combining the survey data from SPF for the Euro Area, we observed that the optimal combinations of more than 50 individual forecasts for inflation and GDP growth performed well compared to the equal-weight combinations used by the ECB. Nevertheless, we observed that these gains were rather modest. A possible explanation is the high similarity of the predictions provided by the different forecasters (see Elliot, 2011).

Although our focus in the paper was restricted to the combination of judgemental point and density forecasts using quadratic loss and logarithmic score, respectively, these schemes can be seen as paradigms for other frameworks such as model averaging and combination (as in Geweke and Amisano, 2011), the use of time-varying weights (as in Billio, Casarin, Ravazzolo and van Dijk, 2011) or of alternative optimization criteria (as in Hansen, 2008). Let us remark that such schemes have also been widely used in other disciplines than economics as e.g. meteorology (Raftery, Gneiting, Balabdaoui and Polakowski, 2005). All these papers consider only a small number of models, and we believe that our results can widen the range of applicability of these more general forecast combination schemes to the situation in which the number of models involved gets large. Such considerations go however beyond the scope of the present paper and are left for further research work.

Acknowlegments

This work was supported by the "Action de recherche concertée" contract ARC-AUWB/2010-15/ULB-11 (CDM and DG) and by the contract GOA-VUB-062 (CDM). Useful discussions with Catherine Dehon, Kajal Lahiri, James Mitchell, Ingmar Nolte, Francesco Ravazzolo and Shaun Vahey are gratefully acknowledged.

References

[1] Amisano, G. and Giacomini, R. (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", *Journal of Business and Economic Statistics*, Vol. 25, 177-190.

[2] Bates, J.M. and Granger, C.W.J. (1969), "The Combination of Forecasts", *Operations Research Quarterly*, Vol. 20, 451-468.

[3] Billio, M., Casarin, R., Ravazzolo, F. and van Dijk, H. K. (2011), "Combining Predictive Densities using Bayesian Filtering with Applications to US Economics Data", preprint.

[4] Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A. and Rautanen, T. (2007), "The ECB survey of professional forecasters (SPF) A review after eight years experience", Occasional Paper Series 59, European Central Bank.

[5] Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A. and Rautanen, T. (2010), "An Evaluation of the Growth and Unemployment Forecasts in the ECB Survey of Professional Forecasters", *Journal of Business Cycle and Measurement Analysis*, Vol. 2, No. 4, 1-28.

[6] Brodie, J., Daubechies, I., De Mol, C., Giannone, D. and Loris, I. (2009), "Sparse and Stable Markowitz Portfolios", *Proceedings of the National Academy of Science*, Vol. 106, No. 30, 12267-12272.

[7] Clark, T.E. and McCracken, M.W. (2010), "Averaging forecasts from VARs with uncertain instabilities", *Journal of Applied Econometrics*, Vol. 25, No. 1, 5-29.

[8] Clemen, R.T. (1989), "Combining Economic Forecasts: A review and annotated bibliography", *International Journal of Forecasting*, Vol. 5, 559-583.

[9] Clements, M.P. and Harvey, D.I. (2007), "Combining Probability Forecasts", *International Journal of Forecasting*, Vol. 27, No. 2, 208-223.

[10] De Miguel, V., Garlappi, L. and Uppal, R. (2009), "Optimal Versus Naive Diversification: How Inefficient is the $\frac{1}{N}$ Portfolio Strategy?", *The Review of Financial Studies*, Vol. 22, No. 5, 1915-1953.

[11] De Mol, C., Giannone, D. and Reichlin, L. (2008), "Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components?", *Journal of Econometrics*, Vol. 146, 318-328

[12] Diebold, F.X., Tay, A.S., and Wallis, K.F. (1999), "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters", in Engle, R.F. and White, H. (eds.), Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger, pp. 76-90. Oxford: Oxford University Press.

[13] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least angle regression", *Annals of Statistics*, Vol. 32, 407-499.

[14] Elliott, G. (2011), "Averaging and the Optimal Combination of Forecasts", mimeo.

[15] Garcia, J.A. (2003), "An introduction to the ECBs survey of professional forecasters", Occasional Paper Series 08, European Central Bank.

[16] Genest, C. and Zidek, J.V. (1986), "Combining Probability Distributions: A critique and an Annotated Bibliography", *Statistical Science*, Vol. 1, No. 1, 114-148.

[17] Genre, V., Kenny, G., Meyer A. and Timmermann, A. (2010), "Combining the Forecasts of the ECB Survey of Professional Forecasters. Can anything beat the Simple Average?", ECB Working Paper No 1277.

[18] Geweke, J. and Amisano, G. (2011), "Optimal Prediction Pools", *Journal of Econometrics*, Vol. 164, 130-141.

[19] Geweke, J. and Amisano, G. (2012), "Prediction with Misspecified Models", *American Economic Review*, Vol. 102, No. 3, 482486.

[20] Giacomini, R. and White, H. (2006)., "Tests of Conditional Predictive Ability," *Econometrica*, Vol. 74, 1545-1578.

[21] Giannone, D., Henry, J., Lalik, M. and Modugno, M. (2010), "An Area-Wide Real-Time database for the Euro Area", *Review of Economics and Statistics*, forthcoming.

[22] Granger, C.W.J. and Ramanathan, R. (1984), "Improved Methods of Combining Forecasts", *Journal of Forecasting*, Vol. 3, 197-204.

[23] Hall, S.G. and Mitchell, J. (2007), "Combining Density Forecasts", *International Journal of Forecasting*, Vol. 23, 1-13.

[24] Hansen, E.B. (2008), "Least-squares forecast averaging", *Journal of Econometrics*, Vol. 146, 342-350.

[25] Jore, A.S., Mitchell, J. and Vahey, S.P. (2010), "Combining forecast densities from VARs with uncertain instabilities", *Journal of Applied Econometrics*, Vol. 25, No. 4, 621-634.

[26] Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005), "Using Bayesian Model Averaging to Calibrate Forecast Ensembles", *Monthly Weather Review*, Vol. 133, 1155-1174.

[27] Sloughter, J.M., Gneiting, T. and Raftery, A.E. (2010), "Probabilistic Wind Speed Forecasting Using Ensemble and Bayesian Model Averaging", *Journal of the American Statistical Assosciation*, Vol. 105, No. 489, 25-35.

[28] Smith, J. and Wallis, K.F. (2009), "A Simple Explanation of the Forecast Combination Puzzle", *Oxford Bulletin of Economics and Statistics*, Vol. 71, No. 3, 331-355.

[29] Stock, J.H. and Watson, M. (2002), "Macroeconomic Forecasting using Diffusion Indexes", *Journal of Business and Economic Statistics*, Vol. 20 No. 2, 147-162.

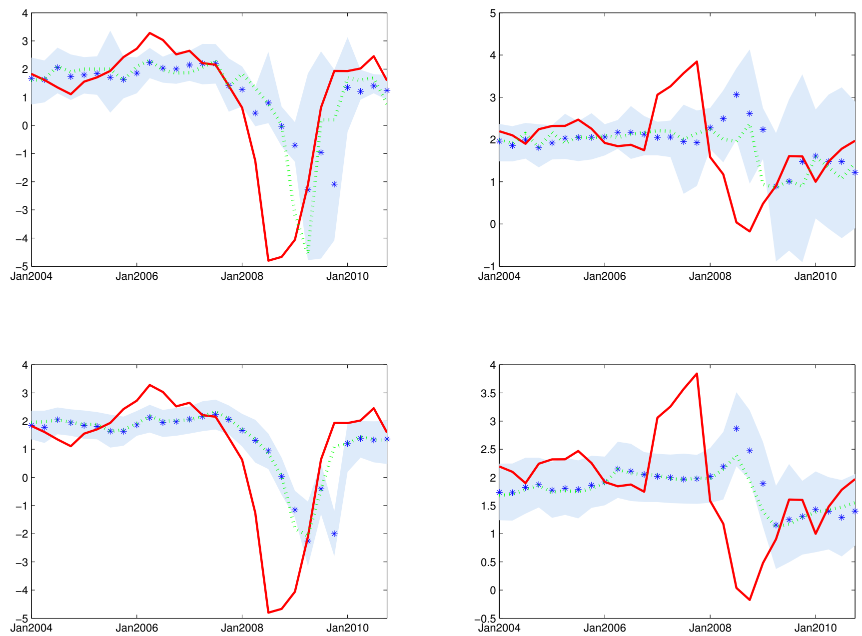
[30] Tay, A.S and Wallis, K.F. (2000), "Density Forecasting: A Survey", *Journal of Forecasting*, Vol. 19, 235-254.

[31] Tibshirani, R. (1996), "Regression shrinkage and selection via the Lasso" *Journal of Royal Statistical Society, Series B*, Vol. 58, 267-288.

[32] Timmermann, A. (2006), "Forecast Combination", Handbook of Economic Forecasting Volume 1 (G. Elliott, C. Granger and A. Timmermann, eds.), North Holland.

[33] Wallis, K.F. (2011), "Combining Forecasts - Forty Years Later", *Applied Financial Economics*, Vol. 21, 33-41.

Figure 1: Density combination



First column: GDP; second column: Inflation. First row: optimal combination; second row: ECB. The blue star line is the median, the shaded area represents the 68%, and the true series is the red solid line. The green dotted line refers to the corresponding optimal point forecasts.