Journal of Econometrics 177 (2013) 213–232

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Time-varying combinations of predictive densities using nonlinear filtering[☆]

Monica Billio[a], Roberto Casarin[a], Francesco Ravazzolo[b], Herman K. van Dijk[c,*]

[a] University of Venice, GRETA Assoc., Italy
[b] Norges Bank and BI Norwegian Business School, Norway
[c] Erasmus University Rotterdam, VU University Amsterdam and Tinbergen Institute, The Netherlands

ARTICLE INFO

Article history:
Available online 17 April 2013

JEL classification:
C11
C15
C53
E37

Keywords:
Density forecast combination
Survey forecast
Bayesian filtering
Sequential Monte Carlo

ABSTRACT

We propose a Bayesian combination approach for multivariate predictive densities which relies upon a distributional state space representation of the combination weights. Several specifications of multivariate time-varying weights are introduced with a particular focus on weight dynamics driven by the past performance of the predictive densities and the use of learning mechanisms. In the proposed approach the model set can be incomplete, meaning that all models can be individually misspecified. A Sequential Monte Carlo method is proposed to approximate the filtering and predictive densities. The combination approach is assessed using statistical and utility-based performance measures for evaluating density forecasts of simulated data, US macroeconomic time series and surveys of stock market prices. Simulation results indicate that, for a set of linear autoregressive models, the combination strategy is successful in selecting, with probability close to one, the true model when the model set is complete and it is able to detect parameter instability when the model set includes the true model that has generated subsamples of data. Also, substantial uncertainty appears in the weights when predictors are similar; residual uncertainty reduces when the model set is complete; and learning reduces this uncertainty. For the macro series we find that incompleteness of the models is relatively large in the 1970's, the beginning of the 1980's and during the recent financial crisis, and lower during the Great Moderation; the predicted probabilities of recession accurately compare with the NBER business cycle dating; model weights have substantial uncertainty attached. With respect to returns of the S&P 500 series, we find that an investment strategy using a combination of predictions from professional forecasters and from a white noise model puts more weight on the white noise model in the beginning of the 1990's and switches to giving more weight to the professional forecasts over time. Information on the complete predictive distribution and not just on some moments turns out to be very important, above all during turbulent times such as the recent financial crisis. More generally, the proposed distributional state space representation offers great flexibility in combining densities.

1. Introduction

When multiple forecasts are available from different models or sources it is possible to combine these in order to make use of all relevant information on the variable to be predicted and, as a consequence, to produce better forecasts. One of the first papers on forecasting with model combinations is Barnard (1963), who considered air passenger data, and see also Roberts (1965) who introduced a distribution which includes the predictions from two experts (or models). This latter distribution is essentially a

☆ We benefited greatly from discussions with Concepción Ausín, Marco Del Negro, Frank Diebold, John Geweke, Dimitris Korobilis, Frank Schorfheide, Xuguang Sheng, and Michael Wiper. We also thank conference and seminar participants at: the 4th CSDA International Conference on Computational and Financial Econometrics, the 6th Eurostat Colloquium, Norges Bank, the NBER Summer Institute 2011 Forecasting, the 2011 European Economic Association and Econometric Society, the Deutsche Bundesbank and Ifo Institute workshop on "Uncertainty and Forecasting in Macroeconomics", Boston University, Federal Reserve Bank of New York, and University of Pennsylvania for constructive comments. HKvD's research is supported by The Netherlands Organisation for Scientific Research (NWO) Secondment Grant, no. 400-07-703. MB and RC's research is supported by the Italian Ministry of Education, University and Research (MIUR) PRIN 2010-11 grant, and by the European Commission Collaborative Project SYRTO. The views expressed in this paper are our own and do not necessarily reflect those of Norges Bank.
* Corresponding author.

E-mail addresses: billio@unive.it (M. Billio), r.casarin@unive.it (R. Casarin), francesco.ravazzolo@norges-bank.no (F. Ravazzolo), hkvandijk@ese.eur.nl (H.K. van Dijk).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.04.009

weighted average of the posterior distributions of two models and is similar to the result of a Bayesian Model Averaging (BMA) procedure. See Hoeting et al. (1999) for a review on BMA, with a historical perspective. Raftery et al. (2005) and Sloughter et al. (2010) extend the BMA framework by introducing a method for obtaining probabilistic forecasts from ensembles in the form of predictive densities and apply it to weather forecasting.

Our paper builds on another stream of literature, starting with Bates and Granger (1969) and dealing with the combination of predictions from different forecasting models; see Granger (2006) for an updated review. Granger and Ramanathan (1984) extend Bates and Granger (1969) and propose to combine forecasts with unrestricted regression coefficients as weights. Liang et al. (2011) derive optimal weights in a similar framework. Billio et al. (2000) and Terui and van Dijk (2002) generalize the least square weights by representing the dynamic forecast combination as a state space with weights that are assumed to follow a random walk process. This approach has been extended by Guidolin and Timmermann (2009), who introduce Markov-switching weights, and by Hoogerheide et al. (2010), who 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 "dynamic model averaging", following the spirit of Terui and van Dijk (2002), and speed up computations by applying forgetting factors in the recursive Kalman filter updating. Hansen (2007) and Hansen (2008) compute optimal weights by maximizing a Mallow criterion. Hall and Mitchell (2007) introduce the Kullback–Leibler divergence as a unified measure for the evaluation and combination of density forecasts and suggest weights that maximize such a distance, see also Geweke and Amisano (2010b). Gneiting and Raftery (2007) recommend strictly proper scoring rules, such as the cumulative rank probability score.

In this paper, we assume that the weights associated with the predictive densities are time-varying and propose a general distributional state space representation of predictive densities and combination schemes. For a review on basic distributional state space representations in the Bayesian literature, see Harrison and West (1997). Our combination method allows for all models to be false and therefore the model set to be misspecified (see Diebold (1991)) or, in other words, incomplete as discussed in Geweke (2010) and Geweke and Amisano (2010b). In this sense we extend the state space representation of Terui and van Dijk (2002) and Hoogerheide et al. (2010) and the model mixing via a mixture of experts (see for example Jordan and Jacobs (1994) and Huerta et al. (2003)). Our approach is general enough to include multivariate linear and Gaussian models (e.g., see Terui and van Dijk (2002)), dynamic mixtures and Markov-switching models (e.g., see Guidolin and Timmermann (2009)), as special cases. 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. Given this general representation, we can estimate (optimal) model weights that minimize the distance between the empirical density and the combination density, by taking into account past performances. In particular, 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 discrete probabilities over the set of predictors. Tests for a specific hypothesis on the values of the weights can be conducted due to their random nature. We discuss weighting schemes with continuous dynamics, which allow for a smooth convex combination of the predictive densities. A learning mechanism is also introduced to enable the dynamics of each weight to be driven by past and current performances of the predictive densities of all models in the combinations.

The constraint that time-varying weights associated with different forecast densities belong to the standard simplex makes the inference process non-trivial and calls for the use of nonlinear filtering methods. We apply simulation based filtering methods, such as Sequential Monte Carlo (SMC), in the context of combining forecasts, see for example Doucet et al. (2001) for a review with applications of this approach and Del Moral (2004) for convergence issues. SMC methods are extremely flexible algorithms that can be applied for inference to both off-line and on-line analysis of nonlinear and non-Gaussian latent variable models used in econometrics. For example, see Billio and Casarin (2010, 2011) for an application to business cycle models and (Creal, 2009) for a review.

Important features of our Bayesian combination approach are analyzed using a set of Monte Carlo simulation experiments. The results are briefly summarized as follows. For the case of a set of linear models, the combination strategy is successful in selecting with probability close to one the true model when the model set is complete. High uncertainty levels in the combination weights appear due to the presence of predictors that are similar in terms of unconditional mean and that differ slightly in terms of unconditional variance. The learning mechanism produces better discrimination between forecast models with the same unconditional mean, but different unconditional variance. The degree of uncertainty in the residuals reduces when the model set is complete. A combination of linear with nonlinear models shows that the learning period may be longer than for the case in which only linear models are present. Finally, we consider an example of a set of models containing a true model with structural instability. The proposed combination approach is able to detect the instability when the model set includes the true model that is generating subsamples of data.

To show practical and operational implications of the proposed approach with real data, this paper focuses on the problem of combining density forecasts using two relevant economic datasets. The first one contains the quarterly series of US real Gross Domestic Product (GDP) and US inflation as measured by the Personal Consumption Expenditures (PCE) deflator. Density forecasts are produced by several of the most commonly used models in macroeconomics. We combine these densities forecasts in a multivariate set-up with model and variable specific weights. For these macro series we find that incompleteness of the models is relatively large in the 1970's, the beginning of the 1980's and during the recent financial crisis while it is lower during the Great Moderation. Furthermore, the predicted probabilities of recession accurately compare with the NBER business cycle dating. Model weights have substantial uncertainty and neglecting it may yield misleading inference on the model's relevance. To the best of our knowledge, there are no other papers applying this general density combination method to macroeconomic data.

The second dataset considers density forecasts on future movements of a stock price index. Recent literature has shown that survey-based forecasts are particularly useful for macroeconomic variables, but there are fewer results for finance. We consider density forecasts generated by financial survey data. More precisely we use the Livingston dataset of six-months-ahead forecasts on the Standard & Poor's 500 (S&P 500), combine the survey-based densities with the densities from a simple benchmark model and provide both statistical and utility-based performance measures of the mixed combination strategy. To be specific, with respect to the returns of the S&P 500 series we find that an investment strategy using a combination of predictions from professional forecasters and from a white noise model puts more weight on the white noise model in the beginning of the 1990's and switches to giving more weight to the professional forecasts over time.

Information on the complete predictive distribution and not just from basic first and second order moments turns out to be