Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 224–237
www.elsevier.com/locate/ijforecast

Forecast combination through dimension reduction techniques

Pilar Poncela[a,*], Julio Rodríguez[a], Rocío Sánchez-Mangas[a], Eva Senra[b,c]

[a] Dept. Análisis Económico: Economía Cuantitativa, Universidad Autónoma de Madrid, Avenida Tomás y Valiente, 5, 28049 Madrid, Spain
[b] Spanish Prime Minister's Economic Bureau, Spain
[c] Dept. Estadística, Estructura Eca. y O.E.I, Universidad de Alcalá, Plaza de la Victoria 2, 28802 Alcalá de Henares, Madrid, Spain

Abstract

This paper considers several methods of producing a single forecast from several individual ones. We compare "standard" but hard to beat combination schemes (such as the average of forecasts at each period, or consensus forecast and OLS-based combination schemes) with more sophisticated alternatives that involve dimension reduction techniques. Specifically, we consider principal components, dynamic factor models, partial least squares and sliced inverse regression.

Our source of forecasts is the Survey of Professional Forecasters, which provides forecasts for the main US macroeconomic aggregates. The forecasting results show that partial least squares, principal component regression and factor analysis have similar performances (better than the usual benchmark models), but sliced inverse regression shows an extreme behavior (performs either very well or very poorly).

Keywords: Combining forecasts; Factor analysis; PLS; Principal components; SIR; Survey of Professional Forecasters

1. Introduction

There is now a growing body of literature on the combination of forecasts, since the early work of Bates and Granger (1969) found that forecast combination may improve forecast accuracy. Surveys on this topic include those of Clemen (1989), Diebold and López (1996), de Menezes, Bunn, and Taylor (2000) and Newbold and Harvey (2002), and, more recently, Timmerman (2006), among others. The recent literature (see for instance Hendry & Clements, 2004; Timmerman, 2006) has stressed the lack of optimality of the individual forecasts (due to mis-specification, mis-estimation, non-stationarities, breaks, partial information, and different loss functions of the forecasters). Forecast combination could improve the performances of individual forecasts, in terms of forecast accuracy. One of the most popular combining techniques is the average of the forecasts at each time t, that is, assigning equal weights to each of the individual forecasts. Surprisingly, on many occasions this easy-to-implement procedure turns out to be hard for

* Corresponding author.
E-mail address: pilar.poncela@uam.es (P. Poncela).

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2010.01.012

most sophisticated alternatives to beat (see, for instance, Stock & Watson, 2004).

To be more precise, let $y_{i,t+1|t}$ be the 1-period-ahead forecast of a certain variable, using information up to time t, given by forecaster i, $i = 1, 2, \ldots, N$; and let $\boldsymbol{y}_{t+1|t} = (y_{1,t+1|t}, \ldots, y_{N,t+1|t})'$ be an N-dimensional vector of 1-step-ahead forecasts at time t. Our purpose is to produce a single combined 1-step-ahead forecast f_t at time t, with information up to time t, from the N initial forecasts; that is,

$$f_t = \boldsymbol{w}_1' \boldsymbol{y}_{t+1|t},$$

where $\boldsymbol{w}_1 = (w_1^1, \ldots, w_N^1)'$ is the weighting vector. A constant could also be added to the previous combining scheme to correct for a possible bias in the combined forecast. The main aim is to reduce the dimension of the problem from N forecasts to just a single one, f_t.

Sometimes, the final combined forecast is built using more than one linear combination of the forecasters. In that case, let $f_{jt} = \boldsymbol{w}_j' \boldsymbol{y}_{t+1|t}$, and let $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{rt})'$ be the first r estimated linear combinations or "prediction factors" coming from the expert opinions. The combining rule for producing a unique 1-step-ahead forecast $\widehat{y}_{t+1|t}$ from all of the sources of information is given by

$$\widehat{y}_{t+1|t} = \widehat{\beta}_0 + \widehat{\beta}_1 f_{1t} + \cdots + \widehat{\beta}_r f_{rt}, \qquad (1)$$

where $\widehat{\beta} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_r)'$ is the ordinary least squares estimate from the regression

$$\pi_t = \beta_0 + \beta_1 f_{1t-1} + \cdots + \beta_r f_{rt-1} + u_t, \qquad (2)$$

with π_t being the observed variable at time t. Notice that $\widehat{y}_{t+1|t}$ is a true ex-ante forecast, since the coefficients β_j are estimated using only information up to time t and do not include any information from the forecasting sample. To summarize, we have to find two types of weights: first, the $\boldsymbol{w}_j$s for forming the linear combinations f_{jt} which will be used as prediction factors; and second, the βs for combining the prediction factors.

We will consider several multivariate dimension reduction techniques for producing the r linear combinations $(f_{1t}, \ldots, f_{rt})$. We divide them into two groups: (1) those that do not use the information available in the variable (π_t) which we are forecasting when forming the prediction factors $(f_{1t}, \ldots, f_{rt})$, that is, to compute the weighting vectors $\boldsymbol{w}_j$, such as principal component and dynamic factor analyses; and (2) those that use the information available in the variable being forecast (π_t) to compute the weighting vectors $\boldsymbol{w}_j$, such as partial least squares and sliced inverse regression. All of the procedures involve two steps. The first step, dimension reduction, produces the regressors through the different multivariate techniques analyzed. The second step produces the final forecast by estimating Eq. (2) by means of OLS. The different procedures differ only in the first step. We would like to stress that Eq. (2) is also estimated in the case where $r = 1$. Nonetheless, in order to allow a "fair" comparison with the average forecast, we also use the bias corrected average of forecasts proposed by Issler and Lima (2009).

In fact, we are interested in analyzing the performances of different combination schemes when they are applied to surveys. We do not know the models (if any) which were used to produce the forecasts, so we are concentrating on the formation of the linear combinations $(f_{1t}, \ldots, f_{rt})$, given the forecasts of the variable. In particular, we will consider the individual forecasts for several US macroeconomic aggregates from the Survey of Professional Forecasters (henceforth SPF).

Several papers have attempted to make partial or related comparisons of some of the above techniques with large data sets. See for instance Boivin and Ng (2006), Heij, Groenen, and van Dijk (2007), Lin and Tsay (2007), Stock and Watson (2002b), and Wang (2008), for some of the most recent ones. All of them use Stock and Watson's (2002b) database. However, none of the previous papers have analyzed the performance of the dimension reduction techniques when the variables being reduced are themselves forecasts of the same variable. This might affect both the number of combined forecasts r used and, of course, their interpretation. While the variables being combined in Stock and Watson's (2002b) database are heterogeneous (for instance, some of them are real economic activity variables, while others are monetary or financial variables, and some others might be classified as prices or inflation related variables), a key feature of the survey data is that all of the variables being combined are alike. All of them are forecasts of the same macroeconomic indicator, giving a particular correlation structure which is different