

Zeng, Jing

**Conference Paper**

## Forecasting Aggregates with Disaggregate Variables: Does boosting help to select the most informative predictors?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Forecasting, No. B16-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Zeng, Jing (2014) : Forecasting Aggregates with Disaggregate Variables: Does boosting help to select the most informative predictors?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Forecasting, No. B16-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100310>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Forecasting Aggregates with Disaggregate Variables: Does boosting help to select the most informative predictors?

Jing Zeng\*

March 1, 2014

## Abstract

Including disaggregate variables or using information extracted from the disaggregate variables into a forecasting model for an economic aggregate may improve the forecasting accuracy. In this paper we suggest to use boosting as a method to select the disaggregate variables which are most helpful in predicting an aggregate of interest. We compare this method with the direct forecast of the aggregate, a forecast which aggregates the disaggregate forecasts and a direct forecast which additionally uses information from factors obtained from the disaggregate components. A recursive pseudo-out-of-sample forecasting experiment for key Euro area macroeconomic variables is conducted. The results suggest that using boosting to select relevant predictors is a viable and competitive approach in forecasting an aggregate.

*Keywords:* aggregation, macroeconomic forecasting, component-wise boosting, factor analysis

*JEL classification:* C22, C43, C52, C53, C82

---

\*University of Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: jing.zeng@uni-konstanz.de

# 1 Introduction

The issue of forecasting aggregates directly or combining disaggregate forecasts has already been discussed and analyzed in an early contribution by Marcellino et al. (2003) who point out that combining disaggregate forecasts outperforms direct forecasting of the Euro area aggregates. Hubrich (2005), however, argues that due to differences in the disaggregate series it is not necessarily better to employ the aggregation of the disaggregate forecasts rather than the direct forecasting method. Even asymptotic theory provides inconclusive results regarding the ranking of these two approaches. In the course of further developments in this discussion, Hendry and Hubrich (2006, 2011) show analytically that taking disaggregate variables into account in the direct forecasting model should be helpful for reducing the forecast mean squared error. Based on the analysis for aggregates with time-varying weights, Lütkepohl (2011) and Brüggemann and Lütkepohl (2013) also point out that incorporating disaggregate information in the forecasting model will generally lead to forecast improvements. However, including all disaggregate variables is often not feasible since too many parameters have to be estimated, which makes this forecast inefficient.

This paper focuses on how to include the disaggregate information or the relevant disaggregate variables in the direct forecasting model. The most promising existing approach of dealing with high-dimensional data, namely, the factor method summarizes the information contained in a large number of series in just a few unobservable common factors. If the factors can be estimated accurately and precisely, then the task of forecasting an aggregate using all disaggregate variables as predictors can be simplified substantially by using the estimated factors as disaggregate information for forecasting. Another way of exploiting high-dimensional data is to select observed raw variables as predictors from a large feasible set which are most informative in predicting an aggregate of interest. This paper suggests to use boosting as such a variable selection device.

Boosting stems from the machine learning and biostatistics literature for analyzing high-dimensional data and has proven to be very competitive in terms of prediction accuracy (Bühlmann and Hothorn (2007)). It estimates

an unknown function iteratively and adds in each iteration the variable with the largest contribution to the fit. Until now, there are only very few applications in the macroeconomic literature. Bai and Ng (2009) estimate the common factors of a large set of predictors and then use boosting to pick out the most relevant factors to be augmented to a standard forecasting model. They find that some form of boosting outperforms the standard factor-augmented forecasts. Buchen and Wohlrabe (2011) evaluate the forecasting performance of boosting and compare it with the forecast combination schemes and dynamic factor models. Using the same data presented in Stock and Watson (2006) they find that boosting is a viable and computationally efficient approach to improve the forecasting accuracy. By investigating the forecasting performance of multivariate models Carriero et al. (2011) also include multivariate boosting in their forecast comparison. They show that boosting performs best in forecasting CPI inflation one month ahead.

This paper compares the performance of four different forecasting models: the direct forecast of an aggregate of interest, aggregating forecasts of disaggregate variables, including the disaggregate information summarized in factor series in the direct forecasting model and using boosting to select relevant disaggregate variables in the direct forecasting model. The main purpose of this study is to answer two empirical questions: First, does adding disaggregate information or variables in the direct forecasting model improve the forecasting performance? Second, what is the appropriate way to incorporate the disaggregate information or variables in the direct forecasting model?

In the empirical application to the Euro area macroeconomic key variables such as real GDP and consumer price index we find that substantial improvements in terms of forecasting accuracy can be achieved when taking the disaggregate variables or information into account in the direct forecast of the aggregate. Furthermore, using boosting to select the most relevant disaggregate variables can produce fairly good forecasts. By determining the optimal number of iterations for boosting, the cross-validation method beats always the data-dependent information criterion.

The structure of the remaining paper is organized as follows. Section 2 outlines the boosting procedure, especially the componentwise boosting. The alternative forecasting models are discussed in Section 3. Section 4 describes

the empirical analysis and Section 5 concludes.

## 2 Componentwise Boosting

### Basis idea

Boosting is a stagewise additive modelling algorithm. The underlying idea is to combine simple estimators to obtain an ensemble such that the performance of each single ensemble member is improved. It estimates an unknown function  $f(x_t)$  for a response variable  $y_t$  as a sum of  $M$  estimated functions:

$$\hat{f}(x_t) = \hat{f}^{(0)} + \nu \sum_{m=1}^M \hat{g}^{(m)}.$$

$x_t = (x_{1,t}, \dots, x_{N,t})'$  contains the  $N$ -dimensional predictor variables which are stationary.  $\hat{f}^{(0)}$  represents the initial function value and the estimated functions  $\hat{g}^{(m)}$  are derived by using a base learner which is a simple fitting procedure based on the minimization of some loss functions. Thus in each iteration  $m$ , boosting adds the estimated base learner  $\hat{g}^{(m)}$  providing the smallest loss. Without modifying the coefficients of those learners which are previously added to the model, only the parameters of the last base learner need to be estimated in each step.

$\nu$  is the shrinkage parameter which ensures that the effect of the base learner is shrunken towards zero in order to reduce its variance and to prevent overfitting. Bühlmann and Hothorn (2007) suggest to use a sufficiently small value since with a small  $\nu$  the forecasting accuracy has been empirically found to be potentially better and almost never worse.

### $L_2$ -Boosting

According to the specification of the response variable  $y_t$ , the base learners and the loss functions, different boosting algorithms were designed in literature. For example, for binary classification the AdaBoost of Freund and Schapire (1996) can be applied. The base procedure is a classifier. If the function of interest  $f(x_t)$  is the conditional mean  $f(x_t) = E(y_t|x_t)$ , the

$L_1$ -Boosting or  $L_2$ -Boosting can be used based on the form of the loss functions. Different to the  $L_1$ -loss, which is not differential at some points, the  $L_2$ -Boosting uses the squared errors as the loss function. It is very useful for regressions especially with many variables as candidate predictors. The loss function of the  $L_2$ -Boosting has the form:

$$L(y_t, f) = \frac{1}{2}(y_t - f)^2.$$

This term is scaled by the factor  $1/2$  so that the negative gradient vector can equale the residuals.

## Componentwise $L_2$ -Boosting

The specification of the base learner  $g^{(m)}$  is of great importance. The  $L_2$ -Boosting chooses the linear least squares as the base learner for linear models. In each iteration the base learner is applied to one candidate variable and thus only one variable will be selected in the sense of ordinary least squares fitting (hence componentwise). The componentwise boosting treats the lags of one predictor as separate predictors so that the variables and lags are selected simultaneously from a large set of candidates for forecasting. To mention that for each iteration the same predictor variable or the lag can be selected.

## Componentwise $L_2$ -Boosting algorithm

Consider a linear forecasting model which has the following form:

$$E(y_{t+h}|y_t, x_t) = c + \sum_{i=1}^p \alpha_i y_{t+1-i} + \sum_{j=1}^N \sum_{i=1}^{p_j} \beta_{j,i} x_{j,t+1-i} = c + \gamma' z_t.$$

$y_t$  represents the one-dimensional target response variable and the vector  $x_t = (x_{1,t}, \dots, x_{N,t})'$  contains the  $N$ -dimensional exogenous predictor variables.  $p$  and  $p_j$  denote the number of the lags for  $y_t$  and  $x_t$ . It is assumed that the variables  $y_t$  and  $x_t$  are stationary.  $z = (y, x_1, \dots, x_N)'$  contains all predictor variables and their lags. So there are totally  $(p + p_1 + \dots + p_N)$  candidate predictors.

1. Start with  $\hat{f}_t^{(0)} = \bar{y}$ . Set  $m = 0$ .
2. Increase  $m$  by 1:
  - Compute the negative gradient of the loss function  $-\frac{\partial L(y_t, f)}{\partial f}$  and evaluate at the estimate of the previous iteration  $\hat{f}_t^{(m-1)}$ :  $u_t = y_t - \hat{f}_t^{(m-1)}$  for  $t = 1, \dots, T$  (current residual).
  - Regress  $u_t$  on each predictor  $z_{(k)}$ ,  $k = 1, \dots, (p + p_1 + \dots + p_N)$  and compute  $SSR_{(k)} = \sum_{t=1}^T (u_t - z_{t,(k)} \hat{\theta}_{(k)})^2$ .
  - Choose the predictor  $z_{(k^*)}$  which minimizes  $SSR$ .
  - Set  $\hat{g}^{(m)} = \hat{\theta}_{(k^*)} z_{(k^*)}$ .
  - Update  $\hat{f}_t^{(m)} = \hat{f}_t^{(m-1)} + \nu \hat{g}^{(m-1)}$ .
  - Iterate step 2 and 3 until  $m = m_{stop}$ .

In the first step of the componentwise boosting, the mean of the target variable  $y_t$  is used as the initial value. Next, the negative gradient of the loss function is computed and this is evaluated at the estimate of the previous iteration. As explained before the squared errors scaled by the factor  $1/2$  are used as the loss function, so we obtain for this step the current residual which is just the difference between the actual data and the fitted value up to that iteration. This current residual is then regressed on the  $(p + p_1 + \dots + p_N)$  candidate variables in turn. The variable which produces the smallest sum of squared errors is selected and can enter the next iteration with its fitted value. The algorithm terminates when the final iteration  $m_{stop}$  is reached. Generally, as stated in Bühlmann and Yu (2003), "  $L_2$ -Boosting is nothing else than repeated least squares fitting of residuals".

## Key parameters

In order to minimize the expected forecasting errors, the bias and variance should be balanced. To reduce the variance it is important to consider a weak learner which involves few parameters and thus has low variance relative to bias. As discussed before, this can be achieved by choosing the shrinkage parameter  $\nu$  sufficiently small. However, a small value of the shrinkage parameter results in a large number of boosting iterations. With a large number

of  $m_{stop}$ , it is clear that the boosting algorithm will over-fit the data. So, how to stop a boosting procedure at an optimal iteration number is of great importance. This key parameter can be determined by using a corrected version of the Akaike information criterion (AIC) or via cross-validation techniques.

The corrected AIC for the iteration step  $m$  has the following form:

$$cAIC(m) = \log(\hat{\sigma}^2) + \frac{1 + df(m)/T}{[1 - df(m) + 2]/T},$$

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{f}_m)^2.$$

$df(m)$  are the degrees of freedom in the boosting fit in the iteration  $m$ . Details are given in Bühlmann and Hothorn (2007).

In the literature it is also advised to use the bootstrapped cross-validation method to choose an appropriate number of iterations. This method determines the optimal  $m_{stop}$  by running the boosting algorithm multiple times for the bootstrapped samples drawn with replacement from the original dataset. Due to the high correlations in the time-series data considered in this paper, the new samples are drawn from the residuals after fitting the model first.

### 3 Forecasting Models

As stated in the introduction, this paper is interested in forecasting an aggregate macroeconomic variable. Suppose the aggregate of interest is  $y_t = w'x_t$  and  $x_t = (x_{1,t}, \dots, x_{N,t})'$  is the vector of disaggregate component variables.  $N$  denotes the number of disaggregate variables.  $w = (w_1, \dots, w_N)'$  is the weighting vector with fixed values. A forecast of  $y_{t+h}$  at period  $t$  is denoted by  $y_{t+h|t}$ .

#### Direct forecast of the aggregate

The direct forecast of the univariate aggregate can be computed as:

$$y_{t+h|t}^{AGG} = c + \sum_{j=1}^p \alpha_j y_{t+1-j}^{AGG} + \epsilon_{t+h}.$$



This forecasting model does not take any disaggregate variables or information into account. The predictors are only the lagged values of the aggregate.

## Aggregation of disaggregate forecasts

In this approach the forecasts are at first computed for each disaggregate variable  $x_1, \dots, x_N$ :

$$x_{i,t+h|t} = c_i + \sum_{j=1}^{p_j} \alpha_{ij} x_{i,t+1-j} + \epsilon_{i,t+h},$$

where  $i = 1, \dots, N$ . Thus  $N$  individual forecasts for  $N$  disaggregate variables can be obtained. Then these forecasts are aggregated using the same weighting vector  $w$  to construct the forecast of  $y_t$ :

$$y_{t+h|t}^{AGG} = w_1 x_{1,t+h|t} + \dots + w_N x_{N,t+h|t}.$$

Since fixed weights are considered here, it is possible to pool the disaggregate forecasts to form a forecast of the aggregate. The case of time-variant aggregation weights (Brüggemann and Lütkepohl (2013)) is not considered in this paper.

## Direct Forecast using disaggregate information summarized in factors

A widely discussed method to extract the information contained in a large number of predictors uses a factor model which has been proposed in Stock and Watson (2002a,b). In their approach, the information contained in the  $N$ -dimensional stationary disaggregate variables  $x_t = (x_{1,t}, \dots, x_{N,t})'$  can be summarized by a small number of  $r$  unobserved common factors  $F_t$  and an idiosyncratic part:

$$x_t = \Lambda F_t + e_t, \quad t = 1, \dots, T,$$

where  $x_t$  is a  $N \times 1$  vector with stationary components,  $\Lambda$  is a  $N \times K$  matrix of factor loadings,  $F_t$  is the  $r \times 1$  vector of common factors and  $e_t$  is an  $N \times 1$  vector of idiosyncratic components.

The forecast for the aggregate variable  $y_t$  by including disaggregate information can be obtained with two steps. In the first step, the disaggregate variables  $x_t = (x_{1,t}, \dots, x_{N,t})'$  are used to estimate the  $r$ -dimensional unobserved factors. Estimation of the factors is done by using a classical static principle components on  $\tilde{x}_t$ , which is obtained by standardizing  $x_t$  to have zero mean and unit variance (Stock and Watson (2002b)). This step gives a  $r \times 1$ -dimensional time series of common factors. It can be shown that under mild conditions the principal components of  $\tilde{x}_t$  are consistent estimators of the true unobservable factors (see e.g. Stock and Watson (2002a) for details). The choice of the number of factors  $r$  may be based on suitable criteria (see e.g. Bai and Ng (2002)). In the second step, one can add the disaggregate information contained in the factors and produce the  $h$ -step-forecast of  $y_t$  as:

$$y_{t+h|t}^{AGG} = c + \sum_{i=1}^p \alpha_i y_{t+1-i} + \sum_{i=1}^r \sum_{j=1}^{p_j} \beta_i \hat{F}_{i,t+1-j} + \epsilon_{t+h}.$$

This forecasting model has the advantage that it considers information both on the aggregate variable and its disaggregate components. If the factors are correctly estimated, the information condensed in all disaggregate variables can be used in the forecasting to improve its performance.

Since the target variable  $y_t^{AGG}$  is aggregated through the disaggregate variables, it may be interesting to check whether using only factors extracted from the disaggregate variables is helpful for improving the forecasting performance, thus the second variant which has the following form can be considered:

$$y_{t+h|t}^{AGG} = c + \sum_{i=1}^r \sum_{j=1}^{p_j} \beta_i \hat{F}_{i,t+1-j} + \epsilon_{t+h}.$$

The lagged values of the aggregate variable are not included in the forecasting model. Only the factors are used as predictors.

## Direct Forecast including disaggregate variables selected by componentwise boosting

Starting point is a forecasting model which has the following form:

$$y_{t+h|t}^{AGG} = c + \sum_{i=1}^p \alpha_i y_{t+1-i}^{AGG} + \sum_{i=1}^N \sum_{j=1}^{p_j} \beta_{i,j} x_{i,t+1-j} + \epsilon_{t+h}.$$

In this variant the lagged values of the aggregate, all the disaggregate variables and their lags are considered. The variables or the lagged values which are most helpful to improve the forecasting performance are selected by using the componentwise boosting algorithm introduced in Section 2.

The second variant considered in this forecasting model focuses only on the disaggregate variables and their lags:

$$y_{t+h|t}^{AGG} = c + \sum_{i=1}^N \sum_{j=1}^{p_j} \beta_{i,j} x_{i,t+1-j} + \epsilon_{t+h}.$$

The componentwise  $L_2$ -Boosting is used to select among all the disaggregate variables the most informative ones.

## Summary of the forecasting models

It is worth briefly summarizing the main characteristics of all the forecasting models under analysis before moving on to the empirical part. Four alternative models are considered in this paper. Each of them aims at forecasting an aggregate variable which is constructed by aggregating the disaggregate components. The first model considers only information from the aggregate variable, while the second model focuses mainly on the disaggregate components. The last two approaches combine both the aggregate and disaggregate information. The factor model approach extracts relevant information from all the disaggregate variables, while the componentwise boosting algorithm allows us to pick out the most relevant observed raw data. Both of them provide a dimension reduction in the predictors.

In the last two forecasting models another variant is considered which performs a dimension reduction only in the disaggregate variables. The reason for this model variant is that the aggregate variables are constructed by

aggregation of the disaggregate variables. When combining the aggregate and disaggregate information together, the information set is used twice. So it may be interesting to investigate whether only using information from the disaggregate variables can provide a better forecasting performance.

## 4 Empirical Analysis

### 4.1 Data

The forecasting comparison considered in this paper includes six macroeconomic key variables for the Euro area on a quarterly frequency: real GDP (YER), the consumer price index (CPI), the GDP deflator (YED), the exchange rate against the US-Dollar (EER) and short- and long-term interest rates (STN and LTN).

The data for aggregated macroeconomic variables are obtained from the Area Wide Model (AWM) database maintained at the Euro Area Business Cycle Network<sup>1</sup>. The AWM aggregates are constructed as the weighted average of log-level data for 11 European countries. The fixed weights used in aggregating the individual country series are the nominal GDP shares for 1995 (Fagan et al. (2001) and Fagan et al. (2005))<sup>2</sup>. This AWM data is now in widespread use, e.g. within the ECB for estimating econometric models.<sup>3</sup> Quarterly data for the period from 1970Q1 to 2011Q4 are used in the following.

The individual member countries' time series data for the same period from 1970Q1 to 2011Q4 are taken from the OECD quarterly national accounts database.<sup>4</sup> Twelve eurozone countries are considered. (Greece, which joined the EMU in 2001, is also included.) All the series are seasonally ad-

---

<sup>1</sup><http://www.eabcn.org>

<sup>2</sup>Weights used in aggregation: Belgium 0.039, Germany 0.305, Spain 0.102, France 0.210, Ireland 0.011, Italy 0.203, Luxembourg 0.002, Netherlands 0.056, Austria 0.030, Portugal 0.024, Finland 0.017. Since Greece, which joined the EMU in 2001 and therefore is not considered in Fagan et al. (2001) and Fagan et al. (2005), has the weight zero in the forecasting comparison.

<sup>3</sup>It should be noted, however, that the AWM database is not an official ECB database.

<sup>4</sup>The data are obtained via Thomson Datastream.

justed and transformed so that they have the same base year. Thus, for each macroeconomic aggregate variable 12 disaggregate European countries data are observed ( $N = 12$ ).

A formal unit root analysis has been conducted and there is evidence that all the time series for all countries can be characterized as  $I(1)$  processes.<sup>5</sup> So the first differences of the variables enter the forecasting models. Logarithms are taken of the real GDP, the CPI, the GDP deflator and the exchange rate EER, while the short- and long-term interest rates are not transformed. To estimate the common factors, the first differences of the variables are standardized to have zero mean and unit variance. Especially for the countries data missing observations are present, so the expectation-maximization (EM) algorithm proposed in Stock and Watson (2002a) is applied to construct a balanced panel.

## 4.2 Comparison methodology

A recursive pseudo-out-of-sample forecasting experiment is conducted for comparing the performance of the forecasting models discussed above. We are interested in forecasting the macroeconomic aggregate  $h$  periods ahead  $y_{t+h|t}^{AGG}$ . Forecasting horizons  $h = 1, 2$  and  $4$  are considered. The forecasting covers the period from 2003Q1 to 2011Q4. The  $h$ -step-ahead projection for constructing the forecasts directly is applied, not the iterated multi-step forecast, since the latter one entails estimating a number of models for the disaggregate variables and the common factors that could erode forecast performance.

To evaluate the forecasting accuracy, the mean squared forecast error (MSFE) is used as loss function, where the forecast error is defined as the difference between the estimated value and its actual value  $e_{t+h} = y_{t+h|t}^{AGG} - y_{t+h}^{AGG}$ . The first forecasting model with a fixed lag length of four is used as benchmark. Each MSFE obtained from other forecasting models (and their variants) will be expressed relative to the MSFE obtained from the first forecasting model (referred to as FA). Thus, if the relative MSFE is less than one, the according model is more precise than the direct univariate forecast

---

<sup>5</sup>The results for the unit root tests are available on request.

with the first model. To simplify the notation, the second forecasting model which combines the disaggregate forecasts together is referred as FD in the following. The third model which uses factors as predictors is referred as FF1 and FF2 for its two different variants. Similarly, we use FB1 and FB2 for the two variants of the forecasting model which applies componentwise boosting algorithm to select the most informative predictors.

For the results reported here, a fixed number of four lags for the autoregressive part and two lags for all disaggregate variables are used. Results using the common data-dependent lag selection information criteria are not reported, since similar results can be obtained by checking the robustness of the results with respect to this choice.

The forecasting variants FF1 and FF2 extract factors by using only the country information on the aggregate variable to be predicted. For instance, when forecasting the Euro area real GDP, the factors are estimated from the dataset that only includes real GDP from the 12 member countries and no other variables are taken into account. The number of factors used in FF1 and FF2 has to be determined. When using the information criteria proposed by Bai and Ng (2002), we found that the maximum number of factors are suggested by all three criteria. Giving the relatively low number of disaggregate variables ( $N = 12$ ), this is not a meaningful choice. Therefore, we decided to explore the cumulative percentage of the variance in the dataset that can be explained by the factors and found that for each variable at least 35% of the total variance in the respective data set can be explained by two factors. So two factors with respectively two lags are used in FF1 and FF2.

For the two variants FB1 and FB2 in the last forecasting model, the shrinkage parameter  $\nu$  is set to the commonly used value of 0.1 (Bühlmann and Hothorn (2007)). The crucial parameter  $m_{stop}$  is determined by the corrected version of AIC or the 25-fold bootstrapped cross-validation (referred to as CV). Both of these two methods are considered in order to analyse to what extent the forecasting accuracy of boosting depends on the choice of the optimal iteration number  $m_{stop}$ .

### 4.3 Empirical results

Results from the forecasting comparison for all the considered macroeconomic variables are presented in Table 1. For each variable and forecasting model variant, we report the MSFE of the respective model variant relative to the MSFE of the forecasting model which uses only the lagged values of the aggregate variable as predictors. Forecasting horizons  $h = 1, 2$  and  $4$  are considered. The forecasting evaluation starts in 2003Q1 and ends in 2011Q4. The smallest MSFE ratio for each variable and forecasting horizon is in bold. This table suggests several conclusions.

First, combining the disaggregate forecasts does not necessary provide more accurate forecasts than the forecasts based on the aggregate variable when forecasting the macroeconomic key variables for the Euro area. For example, when forecasting the growth rate of the real GDP (YER), the relative MSFEs for the forecasting model FD at the horizons  $h = 1$  and  $h = 4$  are larger than one. Similar results hold for the variables YED, EER and STN. Only for the variables CPI and STN, gains can be observed by pooling all the country-specific forecasts together.

Second, overall there is evidence that taking into account disaggregate information in form of factors or selected disaggregate variables can improve the forecasting performance. For the variable YER, forecasting models and variants based on factors or boosting procedure lead in almost all horizons to smaller MSFEs. Only one exception is observed for the model variant FB2 using AIC to determine  $m_{stop}$  ( $h = 2$ ). For the other variables, the use of disaggregate information or variables may not lead to sizable gains in forecasting precision in all cases, however, for some forecasting horizons such gains are always observable.

Third, it is difficult to find an appropriate way to incorporate the disaggregate information or disaggregate variables in the forecasting model. In some cases using factors as predictors tends to have lower MSFEs. However, in many cases the boosting method can indeed improve the forecasting performance compared to the factor approach. Furthermore, in Table 1 the smallest MSFE ratio for each variable and forecasting horizon is in bold. It can be seen that especially for the variables EER and interest rates, using

boosting to select the predictors yields for all horizons the smallest relative MSFEs.

Fourth, there is no compelling evidence that using only the disaggregate information or variables, without considering the lagged values of the aggregate variables in the forecasting model, is helpful to improve the forecasting precision, since the second variant of the forecasting model which uses factors as predictors and the second variant of the boosting procedure perform in most cases worse than the respective first variant. This result indicates that even if the aggregate variables are constructed by aggregating the disaggregate variables, the information hidden in the aggregate variables is important and therefore should be considered for forecasting.

Fifth, it can be found that using the cross-validation method to choose the optimal iteration number always outperforms the use of the corrected version of AIC. In all cases the bootstrapped cross-validation leads to larger forecasting improvements than the AIC.

## 5 Conclusion

The main purpose of this paper is to check whether taking into account disaggregate information or variables is helpful to improve the forecasting accuracy of an aggregate variable and how to include the disaggregate information or the relevant disaggregate variables in the direct forecasting model appropriately. A new variable selection method, the boosting procedure is introduced and discussed. In context of forecasting the six macroeconomic variables for the Euro area, four forecasting models (and their variants) are compared: (1) the direct forecast of an aggregate which uses only the lagged values of the aggregate as predictors, (2) first forecasting the disaggregate variables respectively and then aggregating the disaggregate forecasts, (3) including disaggregate information with the help of the factor model in the direct forecast and (4) including disaggregate variables selected by boosting in the direct forecast.

The empirical results indicate that in comparison to the direct forecasting model which uses only past values of the aggregate variable, using disaggregate information summarized in the factors or using disaggregate variables



selected by boosting as additional predictors in the direct forecasting model can provide more accurate forecasts. Moreover, it can be shown that using boosting to select disaggregate variables beats in most cases the benchmark. Thus boosting is a viable and competitive approach alternative to other methods using a large number of predictors. The number of the iterations is the key parameter for the boosting procedure. Based on our empirical results we suggest to use the cross-validation method to determine this important parameter.

Table 1: MSFEs of different forecasting models relative to the first forecasting model

|          | YER          |              |              | CPI          |              |              |
|----------|--------------|--------------|--------------|--------------|--------------|--------------|
| horizon  | $h = 1$      | $h = 2$      | $h = 3$      | $h = 1$      | $h = 2$      | $h = 3$      |
| FD       | 1.018        | 0.980        | 1.020        | <b>0.968</b> | 0.941        | <b>0.925</b> |
| FF1      | 0.833        | 0.964        | 0.970        | 1.002        | 1.008        | 1.015        |
| FF2      | <b>0.825</b> | 0.975        | 0.993        | 1.008        | 1.019        | 1.085        |
| FB1(AIC) | 0.913        | 0.991        | 0.978        | 0.975        | <b>0.931</b> | 0.968        |
| FB1(CV)  | 0.907        | <b>0.961</b> | <b>0.931</b> | 0.974        | <b>0.931</b> | 0.967        |
| FB2(AIC) | 0.909        | 1.009        | 0.984        | 1.010        | 0.937        | 0.993        |
| FB2(CV)  | 0.907        | 0.963        | 0.942        | 1.010        | 0.937        | 0.993        |

  

|          | YED          |              |              | EER          |              |              |
|----------|--------------|--------------|--------------|--------------|--------------|--------------|
| horizon  | $h = 1$      | $h = 2$      | $h = 3$      | $h = 1$      | $h = 2$      | $h = 3$      |
| FD       | 1.045        | 1.108        | 1.125        | 0.998        | 1.004        | 1.000        |
| FF1      | <b>0.890</b> | <b>0.985</b> | 1.013        | 1.004        | 1.008        | 0.992        |
| FF2      | 0.896        | 1.084        | 1.065        | 1.014        | 0.981        | 0.990        |
| FB1(AIC) | 0.909        | 1.043        | 0.986        | 1.010        | 0.975        | 0.992        |
| FB1(CV)  | 0.908        | 1.012        | <b>0.958</b> | <b>0.955</b> | <b>0.924</b> | <b>0.975</b> |
| FB2(AIC) | 0.990        | 1.031        | 1.038        | 1.005        | 0.975        | 0.989        |
| FB2(CV)  | 0.982        | 1.012        | 1.013        | 0.957        | 0.925        | 0.977        |

  

|          | LTN          |              |              | STN          |              |              |
|----------|--------------|--------------|--------------|--------------|--------------|--------------|
| horizon  | $h = 1$      | $h = 2$      | $h = 3$      | $h = 1$      | $h = 2$      | $h = 3$      |
| FD       | 0.960        | 0.955        | 0.996        | 1.056        | 1.027        | 1.012        |
| FF1      | 1.004        | 1.046        | 1.139        | 1.013        | 1.021        | 0.992        |
| FF2      | 1.053        | 1.055        | 1.122        | 0.998        | 1.031        | 1.023        |
| FB1(AIC) | 0.970        | 1.022        | 1.002        | 1.002        | 1.000        | 1.011        |
| FB1(CV)  | 0.920        | <b>0.934</b> | <b>0.888</b> | <b>0.979</b> | <b>0.991</b> | <b>0.989</b> |
| FB2(AIC) | 0.968        | 1.026        | 1.003        | 1.023        | 1.019        | 1.022        |
| FB2(CV)  | <b>0.907</b> | 0.950        | 0.896        | 1.017        | 1.000        | 0.998        |

## References

- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 71(1), 135–172.
- Bai, J. and S. Ng (2009). Boosting diffusion indices. *Journal of Applied Econometrics* 24, 607–629.
- Bühlmann, P. and T. Hothorn (2007). Boosting algorithms: regularization, prediction and model fitting. *Statistical Science* 22, 477–505.
- Bühlmann, P. and B. Yu (2003). Boosting with the l2 loss: Regression and classification. *Journal of the American Statistical Association* 98, 324–339.
- Brüggemann, R. and H. Lütkepohl (2013). Forecasting contemporaneous aggregates with stochastic aggregation weights. *International Journal of Forecasting* 29(1), 60–68.
- Buchen, T. and K. Wohlrabe (2011). Forecasting with many predictors: Is boosting a viable alternative? *Economic Letters* 113, 16–18.
- Carriero, A., G. Kapetanios, and M. Marcellino (2011). Forecasting large datasets with bayesian reduced rank multivariate models. *Journal of Applied Econometrics* 26(5), 735–761.
- Fagan, G., J. Henry, and R. Mestre (2001). An area-wide model for the euro area. Working Papers 42, European Central Bank.
- Fagan, G., J. Henry, and R. Mestre (2005). An area-wide model for the euro area. *Economic Modelling* 22, 39–59.
- Freund, Y. and R. Schapire (1996). Experiments with a new boosting algorithm. *International Conference on Machine Learning*, 148–156.
- Hendry, D. F. and K. Hubrich (2006). Forecasting economic aggregates by disaggregates. Working Papers 589, European Central Bank.
- Hendry, D. F. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics* 29(2), 216–227.

- Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? *International Journal of Forecasting* 21(1), 119–136.
- Lütkepohl, H. (2011). Forecasting nonlinear aggregates and aggregates with time-varying weights. *Journal of Economics and Statistics* 231(1), 107–133.
- Marcellino, M., J. H. Stock, and M. W. Watson (2003). Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review* 47(1), 1–18.
- Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.
- Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.
- Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 515–554. North Holland, Amsterdam.