Econometric Reviews ›

Volume 29, 2010 - Issue 5-6: The Link Between Statistical Learning Theory and Econometrics: Applications in Econometrics, Finance, and Marketing

Original Articles

Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth

David E. Rapach & Jack K. Strauss

Pages 511-533 | Published online: 15 Sep 2010

Cite this article https://doi.org/10.1080/07474938.2010.481550

Abstract

Forecasting a macroeconomic variable is challenging in an environment with many potential predictors whose predictive ability can vary over time. We compare two approaches to forecasting U.S. employment growth in this type of environment. The first approach applies bootstrap aggregating (bagging) to a general-to-specific procedure based on a general dynamic linear regression model with 30 potential predictors. The second approach considers several methods for combining forecasts from 30 individual autoregressive distributed lag (ARDL) models, where each individual ARDL model contains a potential predictor. We analyze bagging and combination forecasts at multiple horizons over four different out-of-sample periods using a mean square forecast error (MSFE) criterion and forecast encompassing tests. We find that bagging forecasts often deliver the lowest MSFE. Interestingly, we also find that

incorporating information from both bagging and combination forecasts based on principal components often leads to further gains in forecast accuracy.

Keywords:

Bagging Combination forecasts Employment Forecast encompassing Principal components

JEL Classification:

C22 C52 C53 E24

ACKNOWLEDGMENTS

The authors thank two anonymous referees for very helpful comments. The authors also gratefully acknowledge financial support from the Simon Center for Regional Forecasting at Saint Louis University. The results reported in this article were generated using GAUSS 6.1. The GAUSS program files are available at http://pages.slu.edu/faculty/rapachde/ Research.htm.

Notes

Lee and Yang (2006) use bagging techniques to develop binary and quantile forecasts of financial variables.

See Timmermann (2006) for a comprehensive review of forecast combining methods.

For example, employment growth—in the context of the so-called "jobless" recovery from the 2001 U.S. recession—received considerable attention during the 2004 presidential election, arguably more than any other economic variable. Employment growth is also viewed as a key indicator of labor market activity and crucial in Federal Reserve formulations of interest rate policy, and hence forecasts of employment growth are important in ruminations about Fed policy actions.

Note that Inoue and Kilian (2008) also compare bagging forecasts of U.S. inflation to combination forecasts computed from individual models. However, they do not consider methods for combining individual model forecasts based on discount MSFE, clusters, or principal components, and these are the combining methods that perform the best with respect to forecasting U.S. employment growth in Rapach and Strauss (2008).

The t-statistics for the OLS estimates of δ_i in (1) are computed using Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) standard errors based on a lag truncation of $h - 1$.

Inoue and Kilian (2008) consider a range of critical values. We obtain similar results using other conventional critical values such as 1.96.

Following Inoue and Kilian (2008), we use $m = h$ and $B = 100$.

"Recursive" indicates that the forecasts are generated using an expanding estimation window. The out-of-sample forecasts are "simulated"—as opposed to "real-time"—because they are based on revised data and not on the data actually available at the time of forecast formation. Real-time forecasting exercises are not feasible in the present article, as data vintages are not readily available for all of the variables we consider in our forecasting exercise. We follow much of the macroeconomic forecasting literature, including Stock and Watson (1999, 2003, 2004), in analyzing simulated out-of-sample forecasts.

Our results are not very sensitive to the maximum lag lengths.

Observe that the number of clusters serves to define the size of the first cluster, as none of the other clusters are used in generating the forecast. The greater the number of clusters, the smaller the size of the first cluster.

Using the taxonomy in Huang and Lee (2007), all of the combining methods we consider are classified as "combination of forecasts," while the bagging forecasts—which are formed by including all of the variables in a single general model—are classified as "combination of information." Another combination of information procedure that could be used is the Stock and Watson (2002) diffusion index, which involves extracting principal components from the potential predictors themselves (instead of extracting principal components from the individual ARDL forecasts). We experimented with diffusion index forecasts and found that the PC combining method forecasts generally perform better. It would be interesting in future research to consider the approach of Armah and Swanson (2007), who adapt the methodology of Bai and Ng

(2006) to a forecasting environment and select individual predictors to serve as proxies for estimated principal components in a~modified diffusion index forecasting model.

The notion of forecast encompassing is developed in, inter alia, Granger and Newbold (1973) and Chong and Hendry (1986). See Clements and Hendry (1998) for a textbook treatment of forecast encompassing.

A word of caution is in order with respect to the use of the MHLN $_h$ statistic in making inferences on the relative information content across forecasting models. Recent research demonstrates that a number of issues—such as the size of the in-sample period relative to the out-of-sample period, type of estimation window (for example, fixed, rolling, or recursive), and whether the models are nested or non-nested—can affect the asymptotic distribution of the test statistics; see Corradi and Swanson (2006) for an informative review of these issues. We recognize that, strictly speaking, all of the conditions required for the validity of the asymptotic distribution may not be met in our applications, so that inferences based on the MHLN $_h$ statistic serve as a rough guide to statistical significance.

Vendor performance is an index that measures how quickly companies receive deliveries from their suppliers. An increase in the index means that it is taking longer for companies to receive deliveries.

Notes: The entries for the AR benchmark model report the MSFE; the other entries report the MSFE for the ARDL model forecasts indicated on the left to the MSFE for the AR benchmark model forecasts.

Notes to Tables 2–5: The MSFE ratio reports the ratio of the MSFE for the BA model or combination forecasts indicated on the left to the MSFE for the AR benchmark model forecasts. H $_0$: BA encompasses CB (H $_0$: CB encompasses BA) corresponds a test of the null hypothesis that the BA model (combination) forecasts encompass the combination (BA model) forecasts against the one-sided, upper-tail alternative hypothesis that the BA model (combination) forecasts do not encompass the combination (BA model) forecasts; () is the OLS estimate of the weight on the combination (BA model) forecast in the optimal convex combination forecast given by (7); MHLN $_h$ is the test statistic corresponding to the null hypothesis; 0.00 indicates less than 0.005; †, *, ** indicate significance at the 10%, 5%, and 1% levels, respectively.

Nevertheless, all of the MSFE ratios for the combination forecasts are below unity, indicating that they outperform the AR benchmark model. It is also worth noting that the MSFE ratio for the BA forecasts is lower than the MSFE ratios for all of the individual ARDL model forecasts in Table 1 over the 1995:04–2005:03 out-of-sample period at the 3- and 6-month horizons.

Note: The entries in the BA model and PC combination rows report the ratio of the MSFE for these forecasts to the MSFE for the AR benchmark model forecasts. The entries in the BA encompasses PC? (PC encompasses BA?) rows indicate whether the BA model (PC combination) forecasts encompass the PC combination (BA model) forecasts according to the results in Tables 2–5 using a 10% significance level. The entries for the average rows report the ratio of the MSFE for a forecast formed as a simple average of the BA model and PC combination forecasts to the MSFE for the AR benchmark model forecasts.

For the cases where both of the $MHLN_h$ statistics are significant, the weights on the BA and PC forecasts in Tables 2-5 are close to 0.50, so taking the mean of the two forecasts is reasonable. This procedure also avoids having to estimate the weights, making it easier to implement in practice.

We examined the robustness of our results along a number of dimensions and obtained similar results. For example, the results are very similar when we use the AIC instead of the SIC to select the lag lengths in (1) and (2). We also computed combination forecasts for a set of potential predictors that excludes manufacturing capital orders and manufacturing and trade sales, two variables that are available with a one-month lag relative to the other potential predictors (and so are not "coincident" with the other predictors). We again obtain very similar results. The complete results for these robustness checks are available at http://pages.slu.edu/faculty/rapachde/Research.htm.

Forecasting US employment growth using forecast combining methods
Source: Journal of Forecasting
How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation
Source: Journal of the American Statistical Association
Econometric Evaluation of Linear Macro-Economic Models
Source: The Review of Economic Studies
Persistence in forecasting performance and conditional combination strategies
Source: Journal of Econometrics
Evaluating latent and observed factors in macroeconomics and finance