International Journal of Forecasting 33 (2017) 102–120

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecasting market returns: bagging or combining?

Steven J. Jordan [a], Andrew Vivian [b,*], Mark E. Wohar [c,b]

[a] Econometric Solutions, United States
[b] Loughborough University, United Kingdom
[c] University of Nebraska-Omaha, United States

ARTICLE INFO

Keywords:
Return forecasting
Fundamentals
Macro variables
Technical indicators
Emerging markets
Asia
G7
Data mining
Bootstrapping

ABSTRACT

This paper provides a rigorous and detailed analysis of bagging methods, which address both model and parameter uncertainty. We provide a multi-country study of bagging, of which there have been very few to date, that examines out-of-sample forecasts for the G7 and a broad set of Asian countries. We find that bagging generally improves the forecast accuracy and generates economic gains relative to the benchmark when portfolio weight restrictions are applied. Bagging also performs well compared to forecast combinations in this setting. We incorporate data mining critical values for appropriate inference on bagging and combination forecast methods. We provide new evidence that the results for bagging cannot be explained fully by data mining concerns. Finally, the forecasting gains are highest for countries with high trade openness and high FDI. The potentially substantial economic gains could well be operational, given the existence of index funds for most of these countries.

1. Introduction

Most of the *out-of-sample* return forecasting evidence has been found for major developed countries, with a particular focus on US data. In general, the evidence suggests that it is difficult to outperform a simple benchmark consistently. For example, Goyal and Welch (2003) state: "By assuming that the equity premium was 'like it always has been', a trader would have performed at least as well in most of our samples". The majority of the international literature has considered several macro variables and fundamental predictor variables based on dividends and earnings,[1] and provides mixed evidence on the extent of the out-of-sample (OOS) predictability.[2] However, there has been very little prior international evidence on the amalgamation of information from these different predictor variables (the study by Jordan & Vivian, 2011, is a notable exception that considers one simple forecast combination technique). Our paper addresses this topic in detail using the bagging method, which accounts for model uncertainty and parameter instability explicitly.

In the context of stock returns, it is important to control for model uncertainty and parameter instability, given that the predictive ability of individual variables is unstable

* Corresponding author.
E-mail address: a.j.vivian@lboro.ac.uk (A. Vivian).

[1] Bossaerts and Hillion (1999), Giot and Petitjean (2011), and Jordan and Vivian (2011) use fundamental predictors. Rapach, Wohar, and Rangvid (2005) examine macro predictors; and Rapach and Wohar (2009) use the dividend-price ratio.

[2] Jordan (2012) examines one particular type of time-variation in aggregate returns, namely long-term reversals. He finds that these reversals can be explained using asset pricing models with time-varying risk factors and time-varying alphas, which suggests that macroeconomic factors are important determinants of the time variation in equity returns.

http://dx.doi.org/10.1016/j.ijforecast.2016.07.003
0169-2070/

(Bossaerts & Hillion, 1999; Paye & Timmermann, 2006; Rapach & Wohar, 2006; amongst many others). According to Inoue and Kilian (2008, p. 511): "Bagging involves generating a large number of bootstrap resamples of the original forecasting problem, applying a pretest model selection rule to each of the resamples, and averaging the forecasts from the models selected by the pretest on each bootstrap sample". Bagging accounts for the model uncertainty explicitly by allowing the predictive model to change not only over time but also across bootstrap samples for the same time period. Bagging also addresses the issue of parameter instability, by effectively averaging the parameter estimates across bootstrap samples in each time period. Bagging has been implemented recently for the forecasting of economic variables, with applications to US inflation (Inoue & Kilian, 2008) and US employment growth (Rapach & Strauss, 2010); however, to the best of our knowledge, there is very little empirical evidence on the effectiveness of bagging for a dataset of international stock market returns, or in a multi-country setting more broadly.[3]

The main contribution of this paper is to provide a rigorous and detailed analysis of the use of the bagging method (Inoue & Kilian, 2008) for forecasting stock returns in Asia and the G7. The bagging method has also been used for binary prediction problems, where Lee and Yang (2006) considered an asymmetric loss function with an application to forecasting the sign of US stock market returns. The branch of bagging that was examined by Inoue and Kilian (2008) and is followed in this paper applies to linear regression models with continuous variables (predictors and predictand), where it is designed to reduce the mean squared forecast errors. The results of the bagging method relative to the no-predictability benchmark (a random-walk with drift model)[4] are compared to the performances of forecast combination methods, which had received little attention in stock return forecasting applications prior to the recent studies of the US (Rapach, Strauss, & Zhou, 2010) and the G7 countries (Jordan & Vivian, 2011).[5] We extend the literature on bagging in various important dimensions. Firstly, we incorporate the data mining of critical values in order to achieve appropriate inference on bagging (and combination forecast methods), which may be of particular importance for the set of G7 countries. Secondly, we estimate the utility gain that can be derived from bagging continuous variables. Thirdly, we provide comprehensive evidence on bagging from 17 countries, whereas prior studies have considered only a single country; this also enables us to examine cross-country determinants of forecast ability.

First, we investigate whether or not data mining can account for the evidence of predictability that we report. The prior literature has focused on the G7, applying bivariate regressions multiple times. Rapach and Wohar (2006) demonstrated that data mining can account in part for the evidence on aggregate return forecastability in the US. This suggests that data mining could be a concern. The issue of data mining is more acute for the G7 countries, as these countries have been investigated more often than the Asian countries.

Second, in an asset allocation exercise, we investigate whether or not bagging generates utility gains. Earlier empirical investigations of bagging on US macroeconomic series did not examine the economic value of bagging forecasts, while more recent work has done so in single country settings (e.g., see Jordan, Vivian, & Wohar, 2016, for Canadian industries in terms of a sector rotation strategy). From a practical perspective, our results suggest that the implementation of equity index trading strategies based upon the new bagging method in medium and smaller markets could help investors to time-vary their portfolio allocations between debt and equity. Hence, we provide new evidence that bagging can generate economic value that is robust to reasonable trading costs and cannot be accounted for fully by data mining concerns.

Third, we compare aggregate stock return forecasting for 11 Asian countries and the G7 (Japan is in both groups) with previous international studies of equity market return predictability, which generally utilize a similar set of large economies, consisting of the major developed countries.[6] The stock returns of the major developed countries tend to be highly correlated with each other; for example, the monthly correlations amongst country pairs in the G7 are about 0.7. This suggests that data from these countries are only partly independent of each other, and may reflect one (or a small number of) common effect(s).[7] In contrast, the Asian financial markets have much lower correlations with each other, and thus broaden the evidence relative to a sample exclusively of G7 countries. Moreover, the Asian economies are of global interest, given that they produce about 30% of the global economic output, make up over 40% of the world's population, and are (generally) growing rapidly; and, in addition, their financial markets are emerging as an important investment class, despite their

[3] Hillebrand, Lee, and Medeiros (2013) also used bagging to estimate the appropriate restrictions on univariate US stock return predictability regressions, but our setting is rather different, as we are using it in a multivariate predictive setting to tackle model uncertainty in a way analogous to those of Inoue and Kilian (2008) and Rapach and Strauss (2010).

[4] To be consistent with the prior literature, we utilize an historical average benchmark. Robustness results using the AR(p) benchmark provide results that are similar to the historical average.

[5] Forecast combination methods are well established as being effective in improving the forecast accuracy in many disparate applications (Clemen, 1989); however, perhaps surprisingly, they have received little attention for stock returns until recently.

[6] Bossaerts and Hillion (1999) use data from 1969–1995 for the G7 and seven other developed European countries. Rapach et al. (2005) use data from the mid 1970s to the late 1990s for the G7 and five other developed European countries. Rapach and Wohar (2009) study the G7 countries. Giot and Petitjean (2011) use data from the early 1950s until 2005 for 10 developed countries. Jordan and Vivian (2011) use data from 1927–2009 for seven developed countries. McMillan and Wohar (2011) use less sophisticated measures of economic value, but include the G7 countries and four Asian countries. Guidolin, Hyde, McMillan, and Ono (2009) study the G7 countries.

[7] Guidolin et al. (2009) find that the return forecasting performances of the G7 countries could be split into two broad groups: those that are Anglo-Saxon and those that are Continental European. Thus, results that hold for the G7 markets may not transfer well to other markets, meaning that an investigation of countries outside of Anglo-Saxon and Continental Europe is warranted.