International Journal of Forecasting 33 (2017) 745–759

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Predicting recessions with boosted regression trees

Jörg Döpke [a], Ulrich Fritsche [b,1], Christian Pierdzioch [c,*]

[a] University of Applied Sciences, Merseburg, Germany
[b] University Hamburg, Germany
[c] Helmut-Schmidt-University, Hamburg, Germany

ARTICLE INFO

Keywords:
Recession forecasting
Boosting
Regression trees

ABSTRACT

We use a machine-learning approach known as boosted regression trees (BRT) to reexamine the usefulness of selected leading indicators for predicting recessions. We estimate the BRT approach on German data and study the relative importance of the indicators and their marginal effects on the probability of a recession. Our results show that measures of the short-term interest rate and the term spread are important leading indicators. The recession probability is a nonlinear function of these leading indicators. The BRT approach also helps to uncover the way in which the recession probability depends on the interactions between the leading indicators. While the predictive power of the short-term interest rates has declined over time, the term spread and the stock market have gained in importance. The BRT approach shows a better out-of-sample performance than popular probit approaches.

1. Introduction

Against the background of the *Great Recession*, researchers have started to reassess various major linear and nonlinear forecasting approaches (see Bec, Bouabdallah, & Ferrara, 2014; Ferrara, Marcellino, & Mogliani, 2015, among others) and leading indicators that are used widely in applied business-cycle research (see for example Drechsel & Scheufele, 2012). We contribute to this rapidly growing strand of research by using a machine-learning approach known as boosted regression trees (BRT) to reexamine the predictive value of selected leading indicators for forecasting recessions in Germany (on boosting, see Freund & Schapire, 1997; Friedman, 2001, 2002; Friedman, Hastie, & Tibshirani, 2000, or for a survey, see Bühlmann & Hothorn, 2007). The BRT approach is a modeling platform that makes it possible to develop a nuanced view of the relative importance of leading indicators for forecasting recessions, to capture any nonlinearities in the data, and to model interaction effects between leading indicators. The BRT approach combines elements of statistical boosting with techniques studied in the literature on regression trees. Boosting is a machine-learning technique that requires an ensemble of simple base learners to be built and combined in an iterative stagewise process in order to build a potentially complicated function known as a strong learner. The weak learners are simple individual regression trees, and the strong learner results from combining the individual regression trees in an additive way. The ensemble of trees is then used to compute recession forecasts.

Regression trees use recursive binary splits to subdivide the space of leading indicators into non-overlapping regions in order to minimize some loss function. Regression trees lend themselves to the recovery of the infor-

* Correspondence to: Helmut-Schmidt-University, Department of Economics, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany.
E-mail address: c.pierdzioch@hsu-hh.de (C. Pierdzioch).

1 Other affiliations: University Hamburg, KOF ETH Zurich, GWU Research Program on Forecasting, Germany.

http://dx.doi.org/10.1016/j.ijforecast.2017.02.003

mational content of leading indicators because regression trees capture, in a natural way, even complex nonlinearities in the link between the recession probability and leading indicators. Moreover, regression trees are insensitive to the inclusion of irrelevant variables in the list of leading indicators, and are robust to outliers in the data (on regression trees, see Breiman, Friedman, Olshen, & Ston, 1984). Aggregating over regions, and over trees, then allows the recovery of highly complicated links between the recession probability and a leading indicator. In addition, special techniques for the analysis of regression trees have been developed that make it straightforward to trace out the quantitative importance of leading indicators and their marginal effects on the recession probability. Regression trees can also be used to shed light on the ways in which the interaction of leading indicators changes the probability of a recession. While regression trees have several interesting advantages (see Hastie, Tibshirani, & Friedman, 2009, p. 351), their hierarchical structure makes them high-variance predictors. The BRT approach overcomes this drawback by using boosting techniques to combine several regression trees additively in order to form a low-variance predictor.

The BRT approach complements the widely-studied probit approach to the forecasting of recessions, which has been popular in the business-cycle literature since the early 1990s (Estrella & Hardouvelis, 1991; Estrella & Mishkin, 1998), and has also been used extensively as a tool for recession forecasting in recent research (Fritsche & Kuzin, 2005; Proaño & Theobald, 2014; Theobald, 2012, among others). Applications of regression trees to economics can be found in monetary economics (Orphanides & Porter, 2000), empirical finance (Savona, 2014, among others), and political and sports forecasting (see for example Cáceres & Malone, 2013; Lessman, Sung, & Johnson, 2010), as well as in the literature on the determinants of financial crises (see Savona & Vezzoli, 2015, among others). The list of applications of boosting to economics includes the research by Berge (2015), who uses boosting to model exchange rates, and Buchen and Wohlrabe (2011), who compare the performance of boosting with those of other widely-studied forecasting schemes for forecasting the growth rate of U.S. industrial production. Lloyd (2014) and Taieb and Hyndman (2014) use boosting for forecasting the hourly loads of a US utility, and Silva (2014) applies boosting to forecasting wind power generation. Robinzonov, Tutz, and Hothorn (2012) use boosting to forecast the monthly growth rate of German industrial production, and find a good performance at short and medium-term forecast horizons. Lehmann and Wohlrabe (2016) also use boosting to forecast German industrial production. They report various top leading indicators, including turnover, orders, and several survey indicators, for four different forecast horizons. Wohlrabe and Buchen (2014) use boosting to forecast several macroeconomic variables. Bai and Ng (2009) study boosting in the context of factor models. Mittnik, Robinzonov, and Spindler (2015) use boosted regression trees to model the stock market volatility. Closely related to our research is the work of Ng (2014), who uses boosted regression trees to forecast U.S. recessions, and finds that only a few predictors are important for predicting recessions (including interest-rate variables), but also that the relative importance of predictors has changed over time. Ng (2014) does not document marginal effects and abstracts from potential interaction effects of the predictor variables being studied, because the regression trees are restricted to stumps (that is, there is no hierarchical structure of the trees).

We find that measures of the short-term interest rate and the term spread are important leading indicators of recessions in Germany, but also that the BRT approach uses other indicators like business climate indicators and stock market returns to grow trees. The informational content of stock market returns for subsequent recessions is in line with both earlier findings for the U.S. documented by Estrella and Mishkin (1998), and recent theoretical and empirical results reported by Farmer (2012). For the G-7 countries, Bluedorn, Decressin, and Terrones (2013) report that drops in real equity prices are useful predictors of recession starts. Barro and Ursúa (2009) find that stock market crashes help to predict depressions, especially in times of a currency or banking crisis. For Germany, Drechsel and Scheufele (2012) also consider stock market returns as an indicator, but find it insignificant, at least before the *Great Recession*. For earlier evidence of the predictive power of the yield curve, see Duarte, Venetis, and Paya (2005), Estrella, Rodriguez, and Schich (2003), Ivanova, Lahiri, and Seitz (2000), and Rudebusch and Williams (2009), among others. Furthermore, marginal effects reveal nonlinearities in the link between the leading indicators and the probability of being in a recession. For example, the recession probability sharply decreases when the term spread changes sign from negative to positive, and it is a non-linear negatively-sloped function of stock market returns. The recession probability also increases sharply when the short-term interest rate rises above approximately 7%–8%. In contrast, the recession probability is relatively flat for lower short-term interest rates, which implies that monetary policy that operates at the zero-lower bound may have a small effect on the recession probability. At the same time, an investigation of interaction effects shows that the increase in the recession probability that follows an increase of the short-term interest rate is larger in times of a bearish stock market than in times of a bull market. Simulation results and results of an out-of-sample forecasting experiment show that the BRT approach has a better out-of-sample performance as compared to variants of the probit approach.

Section 2 of the paper briefly describes the BRT approach, while Section 3 describes our data. Section 4 then reports our results, and Section 5 concludes.

2. BRT approach

The machine-learning literature has developed several boosting algorithms for solving regression/classification problems under various loss functions (for surveys, see Bühlmann & Hothorn, 2007; Mayr, Binder, Gefeller, & Schmid, 2014a,b; Schapire, 2003). One of the earliest and most popular boosting algorithms is the Adaboost