ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Döpke, Jörg; Fritsche, Ulrich; Pierdzioch, Christian

Working Paper

Predicting recessions in Germany with boosted regression trees

DEP (Socioeconomics) Discussion Papers - Macroeconomics and Finance Series, No. 5/2015

Provided in Cooperation with:
Hamburg University, Department Socioeconomics

Suggested Citation: Döpke, Jörg; Fritsche, Ulrich; Pierdzioch, Christian (2015) : Predicting recessions in Germany with boosted regression trees, DEP (Socioeconomics) Discussion Papers - Macroeconomics and Finance Series, No. 5/2015, Hamburg University, Department Socioeconomics, Hamburg

This Version is available at:
https://hdl.handle.net/10419/148214

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.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Department Socioeconomics

Predicting Recessions in Germany With Boosted Regression Trees.

Jörg Döpke

Ulrich Fritsche

Christian Pierdzioch

DEP (Socioeconomics) Discussion Papers
Macroeconomics and Finance Series
5/2015

Hamburg, 2015

Predicting Recessions in Germany With Boosted Regression Trees

Jörg Döpke, Ulrich Fritsche and Christian Pierdzioch *

University of Applied Sciences Merseburg
University Hamburg
Helmut-Schmidt-University Hamburg

First version: October 2015

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. We then use receiver operating characteristic (ROC) curves to study the accuracy of forecasts. Results show that the short-term interest rate and the term spread are important leading indicators, but also that the stock market has some predictive value. The recession probability is a nonlinear function of these leading indicators. The BRT approach also helps to recover how the recession probability depends on the interactions of 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. We also study how the shape of a forecaster's utility function affects the optimal choice of a cutoff value above which the estimated recession probability should be interpreted as a signal of a recession. The BRT approach shows a competitive out-of-sample performance compared to popular Probit approaches. (188 words)

JEL-Classification: C52, C53, E32, E37
Keywords: Recession forecasting, Boosting, Regression trees, ROC curves

*Corresponding author. Contact: Christian Pierdzioch, Helmut-Schmidt-University, Department of Economics, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany. email: c.pierdzioch@hsu-hh.de.
Other Affiliations: Ulrich Fritsche: University Hamburg, KOF ETH Zurich, GWU Research Program on Forecasting.
Acknowledgments: We thank Artur Tarassow and Christian Proaño for helpful comments.

1 Introduction

Against the background of the *Great Recession*, researchers have started to reassess major forecasting approaches and leading indicators widely used in applied business-cycle research (for example, see Drechsel and Scheufele, 2012). In this research, we use 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 and Schapire (1997); Friedman (2001, 2002); Friedman et al. (2000), for a survey, see Bühlmann and Hothorn (2007)). The BRT approach is a modeling platform that renders 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 building and combining an ensemble of simple base learners in an iterative stagewise process 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 to minimize some loss function that penalizes wrong positive and false negative recession forecasts. Regression trees lend themselves to recover the informational 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 set of leading indicators, and they are robust to outliers in the data (on regression trees, see Breiman et al., 1984). Aggregating over regions, and over trees, then allows highly complicated links between the recession probability and a leading indicator to be recovered. In addition, special techniques have been developed for the analysis of regression trees that make it straightforward to trace out the quantitative importance of a leading indicator for forecasting recessions and the marginal effect of a movement in a leading indicator on the recession probability. Regression trees can also be used to shed light on how the interaction of leading indicators changes the probability of a recession. While regression trees have several interesting advantages (see Hastie et al., 2009, page 351), their hierarchical structure makes them high-variance predictors. The BRT approach overcomes this drawback by using boosting techniques to additively combine several regression trees to form a low-variance predictor.

The BRT approach complements the widely-studied Probit approach to forecasting recessions, which has been popular in the business-cycle literature since the early nineties (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998), and has been extensively used as a tool for recession forecasting also in recent research (Fritsche and Kuzin, 2005; Theobald, 2012; Proaño and Theobald, 2014, among others). Applications of regression trees in economics can be found in monetary economics (Orphanides and Porter, 2000), empirical finance (Savona, 2014; Malliaris and Malliaris, 2015, among others)), and in the literature on the determinants of financial crises (Manasse and Roubini, 2009; Duttagupta and Cashin, 2011; Savona and Vezzoli, 2015, among others). The list of applications

of boosting in economics includes the research by Berge (2015), who uses boosting to model exchange rates, and Buchen and Wohlrabe (2011), who compare how boosting performs relative to other widely studied forecasting schemes with respect to forecasting the growth rate of U.S. industrial production. Robinzonov et al. (2012) use boosting to forecast the monthly growth rate of German industrial production and find a good performance in short-term and medium-term 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 et al. (2015) use boosted regression trees to model stock-market volatility. Closely related to our research is the work by Ng (2014), who uses boosted regression trees to forecast U.S. recessions. Key findings are that only 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, because the regression trees are restricted to stumps (that is, there is no hierarchical structure of the trees), abstracts from potential interaction effects of the predictor variables being studied.

We find that the short-term interest rate and the term spread are top 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 short-term interest rates and stock market returns for subsequent recessions is in line with earlier findings for the U.S. documented by Estrella and Mishkin (1998), and with recent theoretical and empirical results reported by Farmer (2012). For Germany, Drechsel and Scheufele (2012) find that stock market returns are a useful (albeit insginificant) indicator at least before the *Great Recession*. For earlier evidence of the predictive power of the yield curve, see Estrella et al. (2003) and Rudebusch and Williams (2009), among others. Marginal effects further 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 nonlinear negatively-shaped function of stock market returns. The recession probability also sharply increases when the short-term interest rate rises above approximately 6%−8%. By 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 illustrates that the effect of a variation of the short-term interest rate on the recession probability depends to a non-ligible extent on the state of the stock market. The increase in the recession probability that follows an increase of the short-term interest rate is larger in times of a bear market than in times of a bull market. Results derived using receiver operating characteristics (ROC) curves show that the BRT approach has a competitive out-of-sample performance in comparison to variants of the Probit approach (for recent applications of ROC curves in economics, see Berge and Jordà, 2011; Lahiri and Wang, 2013; Liu and Moench, 2014; Savona and Vezzoli, 2015, among others). Moreover, the BRT approach can be combined with ROC curves to study, for forecasters with a symmetric and an asymmetric loss function, the optimal choice of a cutoff value above which the estimated recession probability should be interpreted as a signal of a recession. In sum, while the BRT approach does not change the fact that recession forecasting re-

mains a challenging task, the approach is a useful technique that complements the widely-studied Probit approach.

We organize the remainder of this research as follows. In Section 2, we describe the empirical methods that we use to compute and evaluate forecasts. In Section 3, we describe our data. In Section 4, we describe our results. In Section 5, we conclude.

2 Empirical Methods

2.1 Computation of Forecasts

We use a boosting algorithm to forecast recessions. Studying boosting algorithms has a long tradition in the machine-learning literature where several algorithms have been developed to solve regression and classification problems under various loss functions (for surveys, see Schapire, 2003; Bühlmann and Hothorn, 2007; Mayr et al., 2014a,b). Among the earliest and most popular boosting algorithms is the Adaboost algorithm developed by Freund and Schapire (1997). It received much attention in the machine-learning literature because of its empirical success as a classifier. In later research Friedman et al. (2000) developed variants of the Adaboost algorithm and traced out its links to Logistic regression models. Friedman (2001, 2002) then showed that boosting can be interpreted as a function-approximation problem, and that a gradient descent paradigm, which also applies to regression trees in a straightforward way, can be used to solve this problem in a forward stage-wise additive way

Drawing on the results derived by Friedman et al. (2000), we model recessions as a binary variable, $y_{t+k} \in \{0, 1\}$, where $y_t = 1$ denotes a recession, $t = 1, ...,$ denotes a time index, and k denotes a forecast horizon. The plan is to model the links between recessions and the leading indicators, $\mathbf{x_t} = (x_{t,1}, x_{t,2}, ...)$, by means of a function, $F(\mathbf{x_t})$, so as to minimize the expected value of a loss function, $\mathcal{L}$. Friedman et al. (2000) suggest to choose $F(\mathbf{x_t})$ to minimize the exponential loss function

$$\mathcal{L}(F) = E \exp(-\tilde{y}_{t+k} F(\mathbf{x_t})), \tag{1}$$

where it is common to define, for ease of notation, $\tilde{y}_{t+k} = 2y_{t+k} - 1$, such that $\tilde{y}_{t+k} \in \{-1, 1\}$, and E denotes the conditional expectations operator (that is, expectations are conditional on the leading indicators, $\mathbf{x_t}$). The loss function given in Equation (1) increases when $\tilde{y}_{t+k}$ and $F(\mathbf{x_t})$ have different signs, and it decreases when $\tilde{y}_{t+k}$ and $F(\mathbf{x_t})$ have the same sign. In other words, the loss function decreases when $F(\mathbf{x_t})$ helps to classify recessions at forecast horizon k. Upon denoting the conditional probability of being in a recession as $P(\tilde{y}_{t+k} = 1|\mathbf{x_t})$, the right-hand-side of Equation (1) can be expanded to give

$$\begin{aligned} E \exp(-\tilde{y}_{t+k} F(\mathbf{x_t})) = P(\tilde{y}_{t+k} = 1|\mathbf{x_t}) \exp(-F(\mathbf{x_t})) \\ + P(\tilde{y}_{t+k} = -1|\mathbf{x_t}) \exp(-F(\mathbf{x_t})). \end{aligned} \tag{2}$$

The first-order condition for a minimum can be easily derived as (Friedman

et al., 2000, page 345)

$$\frac{\partial E \exp(-\tilde{y}_{t+k} F(\mathbf{x_t}))}{\partial F(\mathbf{x_t})} = -P(\tilde{y}_{t+k} = 1|\mathbf{x_t}) \exp(-F(\mathbf{x_t})) + P(\tilde{y}_{t+k} = -1|\mathbf{x_t}) \exp(-F(\mathbf{x_t})) \stackrel{!}{=} 0, \quad (3)$$

so that $\mathcal{L}(F)$ is minimized by setting $F(\mathbf{x_t})$ to one-half of the log-odds ratio:

$$F(\mathbf{x_t}) = \frac{1}{2} \log \frac{P(\tilde{y}_{t+k} = 1|\mathbf{x_t})}{P(\tilde{y}_{t+k} = -1|\mathbf{x_t})}. \quad (4)$$

Rearranging terms and using the fact that $P(\tilde{y}_{t+k} = 1|\mathbf{x_t}) = 1 - P(\tilde{y}_{t+k} = -1|\mathbf{x_t})$ yields

$$P(\tilde{y}_{t+k} = 1|\mathbf{x_t}) = \frac{\exp(F(\mathbf{x_t}))}{\exp(-F(\mathbf{x_t})) + \exp(F(\mathbf{x_t}))}, \quad (5)$$

$$P(\tilde{y}_{t+k} = -1|\mathbf{x_t}) = \frac{\exp(-F(\mathbf{x_t}))}{\exp(-F(\mathbf{x_t})) + \exp(F(\mathbf{x_t}))}. \quad (6)$$

The unknown function, $F(\mathbf{x_t})$, that links recessions to the leading indicators, thus, can be estimated by modeling the log-odds ratio. A natural point of departure is the unconditional recession probability (that is, the proportion of times that we observe $y_{t+k} = 1$). The unconditional recession probability, however, is a crude measure of the conditional recession probability, and boosting techniques show how to refine this measure.

The basic idea of boosting is to break down a potentially complicated function estimation problem into a series of simple problems by stipulating that the function to be estimated, $F(\mathbf{x_t})$, can be expressed as the sum of much simpler functions, $T(\mathbf{x_t})$. We have

$$F(\mathbf{x_t}) = \sum_{m=0}^{M} T_m(\mathbf{x_t}), \quad (7)$$

where m is the index of a weak learner, and M denotes some upper bound on the number of functions, $T(\mathbf{x_t})$, that we want to consider in the summation operation. In the machine-learning literature, the complicated function, $F(\mathbf{x_t})$, that we plan to use to forecast recessions is known as a strong learner, and the simpler functions, $T(\mathbf{x_t})$, that we use to approximate this complicated function are known as weak learners.

Rather than finding M distinct weak learners in a single step, boosting algorithms render it possible to estimate the weak learners step-by-step in a forward stage-wise manner. In our research, we use a boosting algorithm known as gradient-descent boosting. The following algorithm describes how gradient-descent boosting works (Friedman, 2001, 2002):

1. Initialize the algorithm: $F_0 = T_0 = \frac{1}{2} \log \frac{P(\tilde{y}_{t+k}=1)}{P(\tilde{y}_{t+k}=-1)}$.
2. Define some upper bound, M, for the number of weak learners.
3. For m in 1 to M:

(a) Compute the negative gradient vector given by $z_{t,m} = -\partial\mathcal{L}(F)/\partial F = \tilde{y}_{t+k} \exp(-\tilde{y}_{t+k} F(\mathbf{x_t}))$.

(b) Fit a weak learner, $T_m(\mathbf{x_t})$, to the negative gradient vector.

(c) Update the function estimate, $F_m(\mathbf{x}_t)$, by adding to $F_{m-1}(\mathbf{x}_t)$ the weak learner, $T_m(\mathbf{x_t})$, as described in Equation (7).

(d) Equipped with the new function estimate, go back to Step (a).

4. When the recursion reaches $m = M$, the strong learner, $F_M(\mathbf{x}_t)$ has been computed as the sum of the weak learners, $T_m(\mathbf{x_t})$, where $m = 0, ..., M$.

In every iteration, the negative gradient vector indicates in which direction to search for a minimum of the loss function. The negative gradient vector is large in absolute value for those observations for which the recession forecast is wrong (false signal or a false nonsignal). The algorithm, thus, uses, in every iteration, a new weak learner so as to shrink what previous iterations have left unexplained, where the weak learners estimated in earlier steps are left unchanged.

In order to make the gradient-descent-boosting algorithm work, we must define what exactly is a weak learner. We use regression trees as weak learners (for nontechnical introductions, see Leathwick et al. (2006); Strobl et al. (2009) among others, see also the textbook by Hastie et al. (2009)). A regression tree, $T(\mathbf{x_t})$, with J terminal nodes partitions in a binary and hierarchical top-down way the space of the leading indicators, $\mathbf{x_t}$, into l non-overlapping rectangular regions, R_l, and predicts, at every terminal node, a region-specific constant, $E(z_{t,m}|\mathbf{x_t} \in R_l)$, of the negative gradient vector. Every region is defined by a leading indicator, $s \in \mathbf{x_t}$, used for the partitioning, and the partitioning point, p (that is, the realization of the leading indicators at which a split is invoked). The splitting indicator and the corresponding partitioning point are chosen to minimize a quadratic loss function defined over the negative gradient vector and the region-specific constant prediction. This problem can be solved by searching over all leading indicators and potential split points. For example, at the top level of a regression tree, the solution to the search problem gives two half-planes, $R_1(s, p)$ and $R_2(s, p)$. The two half-planes give a new tree that now has two terminal nodes. For this new tree, the search is repeated separately for the two half-planes, $R_1(s, p)$ and $R_2(s, p)$, identified at the top level. The new tree then has four nodes. A hierarchical regression tree emerges as this process continues until a maximal tree size defined by the researcher is reached or a minimum number of observations is assigned to the terminal nodes.

The regression tree is integrated into Step 3a of the gradient-descent-boosting algorithm by choosing optimal terminal node responses for the given loss function such that (Friedman, 2002, Algorithm 1)

$$\gamma_{l,m} = \arg\min_{\gamma} \sum_{\mathbf{x_t} \in R_{l,m}} \mathcal{L}(F_{m-1}(\mathbf{x_t}) + \gamma). \tag{8}$$

The minimization problem specified in Equation (8) can be solved using Newton's method (Friedman et al., 2000, page 353). Equation (7) can then be rewritten as

$$F(\mathbf{x_t}) = \sum_{m=0}^{M} \gamma_{l,m} \mathbf{1}_{\mathbf{x_t} \in R_{l,m}} \tag{9}$$

where **1** denotes the indicator function. In other words, the conditional log-odds ratio is estimated by additively combining a large number of trees and the corresponding region-specific terminal nodes.

In order to prevent the algorithm from overfitting, Friedman (2001) introduces a shrinkage parameter, $0 < \lambda \leq 1$, that curbs the influence of individual weak learners on the strong learner. The shrinkage parameter is incorporated into the gradient-descent-boosting algorithm by modifying Step 3c to

$$F_m(\mathbf{x_t}) = F_{m-1}(\mathbf{x_t}) + \lambda \gamma_{l,m} \mathbf{1}_{\mathbf{x_t} \in R_{l,m}}. \tag{10}$$

The shrinkage parameter, thus, can be interpreted as a learning rate that governs by how much a weak learner estimated in iteration m contributes to the function estimate. In the boosting literature, the learning rate typically is set to 0.1 or some smaller value. Fixing the learning rate at a smaller value implies that more iterations, M, are needed to optimize the performance of a forecasting model.

The performance of a forecasting model is the result of a bias-variance trade off (for textbook expositions, see Hastie et al., 2009; James et al., 2013). Estimating more weak learners results in a reduction of bias because a forecasting model more closely tracks even complicated links between the leading indicators and the probability of being in a recession. At the same time, variance increases because the resulting more complex forecasting model (too) closely captures the specific features of the training dataset used to fit the model. Estimating only a few weak learners, in contrast, inflates bias, but also makes overfitting less likely. The resulting bias-variance trade-off implies that it is possible to detect an optimal number, M^*, of base learners.

Building on earlier research on bagging Breiman (1996) and random forests Breiman (2001), Friedman (2002) suggests that injecting randomness at Step 3a of the gradient-descent-boosting algorithm is a further modeling element that helps to improve model performance. The resulting stochastic gradient-descent-boosting algorithm requires that a researcher samples without replacement, before fitting a weak learner, a subset from the data. Only the sampled data are then used to estimate the next weak learner. Adding this element of randomness to the boosting algorithm helps to stabilize model predictions by lowering the correlations of predictions from individual weak learners.

2.2 Evaluation of Forecasts

Equipped with an estimate of the function, $F(\mathbf{x_t})$, an estimate, $\hat{P}(y_{t+k} = 1|\mathbf{x_t})$, of the conditional probability of being in a recession can be computed using Equation (5). The estimated conditional recession probability, in turn, can be mapped back into the binary recession classifier by checking $\text{sign}(F(\mathbf{x_t}))$, which implies a recession is predicted whenever $\hat{P}(y_{t+k} = 1|\mathbf{x_t}) > 0.5$. It is all but clear, however, that simply fixing the cutoff value at 0.5 optimally balances the trade-off between forecast errors (see also Ng (2014), page 26; see also the discussion in the textbook by Greene (2003), page 685). Two types of errors can arise: A predicted recession does not occur (false positive), and a recession occurs but was not predicted (false negative). While predicting many recessions helps to maximize the proportion of true positives, a trade-off arises because the proportion of false positives will rise as well. Conversely, if a forecaster only occasionally predicts a recession the proportion of true negatives is high, but

at the same time such a forecasting strategy increases the proportion of false negatives.

Simply opting for another arbitrary cutoff value does not solve the problem either because any shift in the cutoff value only changes the balance between false positives and false negatives. Hence, a more general approach is needed that accounts for a full range of cutoff values and, at the same time, maps the balance between different types of forecast errors into a simple decision criterion that can be used in applied business-cycle research to assess the informational content of estimated recession probabilities. Receiver operating characteristic (ROC) curves are such a general approach to evaluate the predictive power of estimated recession probabilities and, more general, of any leading indicator of the business cycle. While ROC curves are often used in the machine-learning literature to study the predictive power of regression and classification techniques (see, for example, the introductory textbook by James et al., 2013) they have become popular in economics only recently (Berge and Jordà, 2011; Lahiri and Wang, 2013; Liu and Moench, 2014; Savona and Vezzoli, 2015; Pierdzioch and Rülke, 2015), so that we briefly describe how to construct an ROC curve.

The starting point for constructing an ROC curve is to formalize the link between the predicted state of the economy, $\hat{y}_t$, and the estimated conditional recession probabilities as follows

$$\hat{y}_{t+k}(c) = \begin{cases} 1, & \text{if} \quad \hat{P}(y_{t+k} = 1|\mathbf{x_t}) \geq c, \\ 0, & \text{if} \quad \hat{P}(y_{t+k} = 0|\mathbf{x_t}) < c. \end{cases} \tag{11}$$

where c denotes some cutoff value. The domain of the cutoff value is the interval from zero to unity. Alternative choices of the cutoff value result in different frequencies of signals, $\hat{y}_{t+k}(c) = 1$, of recessions, $y_{t+k} = 1$, and an ROC curve summarizes how alternative choices of the cutoff value lead to different proportions of true positives, $PTP(c)$, and true negatives, $PTN(c)$. The proportion of true positives is also known as the *sensitivity* of forecasts, and the proportion of true negatives is known as the *specificity* of forecasts. Sensitivity measures the ratio of true positives relative to all recessions, and specificity measures the ratio of true negatives relative to all nonrecessions. Sensitivity and specificity are both functions of the cutoff-value, c, and are defined as

$$PTP(c) = \frac{1}{n_R} \sum_{t=1}^{N} \mathbf{1}_{y_{t+k}=\hat{y}_{t+k}=1} \tag{12}$$

$$PTN(c) = \frac{1}{n_{NR}} \sum_{t=1}^{N} \mathbf{1}_{y_{t+k}=\hat{y}_{t+k}=0} \tag{13}$$

where n_R denotes the number of recession periods (true positives plus false negatives) and n_{NR} denotes the number of periods that the economy was not in a recession (true negatives plus false positives), where $N = n_R + n_{NR}$.

If the cutoff value reaches its upper bound a forecasting model never produces signals, so that there are no true positives and no false positives. For such a cutoff value, it follows that $PTP(c) = 0$ and $PTN(c) = 1$, such that $1 - PTN(c) = 0$. Conversely, if the the cutoff value reaches its lower bound a forecasting model always produces signals, implying that there are no false negatives, but many false positives. We have $PTP(c) = 1$ and $1 - PTN(c) = 1$.

Different combinations of the proportion of true positives, $PTP(c)$, and the proportion of false positives, $1 - PTN(c)$, thus can be plotted by varying the cutoff value, and a unit quadrant suffices to draw the resulting ROC curve. A ROC curve shows, in a unit quadrant, combinations of the proportion of false positives on the horizontal axis and the proportion of true positives on the vertical axis. Beginning with a cutoff value fixed at its upper bound, an ROC curve starts at the point $[0, 0]$. The ROC curve then monotonically increases as the cutoff value decreases, and eventually the ROC curve approaches the point $[1, 1]$ as the cutoff value reaches its lower bound.

If recession forecasts are indistinguishable from pure noise forecasts for any given cutoff value, such that always $PTP(c) = 1 - PTN(c)$, then the resulting ROC curve coincides with the bisecting line in a unit quadrant. If recession forecasts dominate a pure noise signal irrespective of the cutoff value then a ROC curve settles above the bisecting line in the north-western part of a unit quadrant because the rate of true positives should always exceed (except at $[0, 0]$ and $[1, 1]$) the proportion of false positives. More informative forecasts result in ROC curves that move deeper into the the north-western part of a unit quadrant. If recession forecasts perform worse than pure noise forecasts for any cutoff value, then the resulting ROC curve settles below the bisecting line in the south-eastern part of a unit quadrant. In this case, however, reversing the definition of a signal again results in a ROC curve that lies above the bisection line. Finally, if recession forecasts perform better than pure noise forecasts for some values of the cutoff value but not for others then a ROC curve emerges that crosses the bisecting line.

The area, AUROC, under a ROC curve is a summary statistic of the performance of recession forecasts. Perfect forecasts result in AUROC = 1 because the corresponding ROC curve hugs to the top left corner of a unit quadrant. Forecasts that are indistinguishable from pure noise forecasts result in AUROC = 0.5 because the ROC curve coincides with the bisecting line. Finally, if recession forecasts systematically perform worse than pure noise forecasts then AUROC < 0.5. In this case, reversing the definition of a signal implies that again AUROC > 0.5.

The AUROC statistic can be estimated using a non-parametric approach that makes use of the result that the AUROC statistic is linked to the Wilcoxon-Mann-Whitney U statistic (Bamber, 1975; Hanley and McNeil, 1982). Like (Greiner et al., 2000, page 38−39), we compute the AUROC statistic as follows:

$$AUROC = \frac{n_{NR}n_R - U}{n_{NR}n_R}, \tag{14}$$

where $U = R - 0.5n_{NR}(n_{NR} + n_R)$ denotes the two-sample Mann-Whitney rank-sum test, and R denotes the rank sum of the nonevents.

We compute confidence bands for the AUROC statistic either by means of simulations or using the formula for its standard error, σ_{AUR}, given by Hanley and McNeil (1982) and Greiner et al. (2000) as

$$\sigma_{AUR} = \sqrt{\frac{A + B + C}{n_{NR}n_R}}, \tag{15}$$

where $A = AUROC(1 - AUROC)$, $B = (n_R - 1)(Q_1 - AUROC^2)$, $C =$

$(n_{NR} - 1)(Q_2 - AUROC^2)$, with $Q_1 = AUROC/(2 - AUROC)$ and $Q_2 = 2AUROC^2/(1 + AUROC)$.

3 Data

3.1 Defining Recessions

Defining a benchmark for recession periods is difficult as what constitutes a "recession" is controversial. Most researchers who study U.S. data refer to the understanding of a recession as coined by the NBER Business Cycle Dating committee National Bureau of Economic Research (2014): "During a recession, a significant decline in economic activity spreads across the economy and can last from a few months to more than a year. Similarly, during an expansion, economic activity rises substantially, spreads across the economy, and usually lasts for several years." In determining recession periods, the committee makes a broad assessment of the overall state of the economy, an approach which is in the tradition of Burns and Mitchell (1946). This so-called "classical" understanding of business cycles is in contrast to the alternative understanding of business cycles as deviations from a long-term trend or potential GDP. In applied research, some efforts have been made to find algorithms that mimic the decisions of the NBER dating committee for U.S. data (Bry and Boschan, 1971; Harding and Pagan, 2003). Such algorithms render it possible to apply the NBER concept also to other countries than the United States. The OECD uses a modified version of the Bry and Boschan (1971) approach as well. Due to its political importance, and to make our results comparable to reference papers using U.S. data, we define recessions in the "classical" sense. Because no "official" business cycle dating committee exists for Germany, throughout our analysis we refer to the dating of peaks and troughs as published by the Economic Cycle Research Institute (2013). Figure 1 shows the industrial production index (5 months centered moving average) for Germany with the recession periods (shaded areas) according to the ECRI definition. The ECRI recessions periods nearly perfectly track the swings in the production cycle and are in line with earlier results reported by Artis et al. (1997), Kholodilin (2005), and Schirwitz (2009), and OECD (2015).

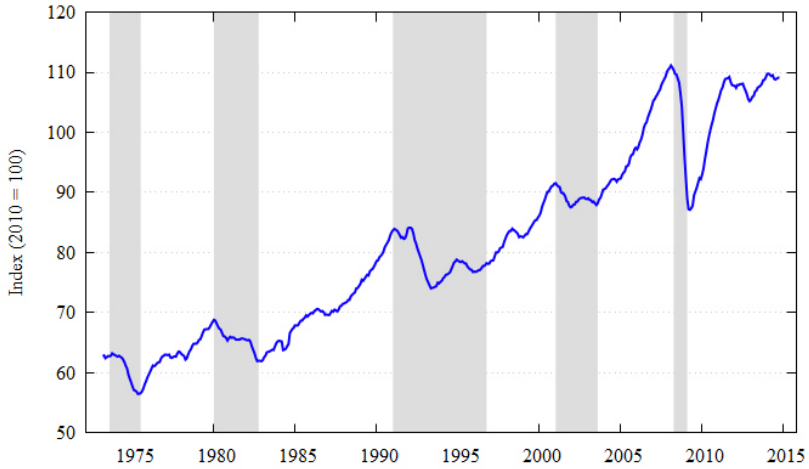
3.2 Selection of Leading Indicators

In earlier literature (Lindlbauer, 1997; Fritsche, 1999; Fritsche and Stephan, 2002; Drechsel and Scheufele, 2012), various time series have been suggested as leading indicators of the German business cycle in general and German recessions in particular (Fritsche and Kuzin, 2005; Theobald, 2012). Like Drechsel and Scheufele (2012), we select leading indicators from the following groups of indicators: (i) financial indicators, (ii) surveys, (iii) real economy variables, (iv) prices, and (v) composite leading indicators. Table 1 lays out the indicators that we study in this research.

We select leading indicators for the following reasons:

1. *Long-term availability:* As recessions are rare events, we need indicators that are available for long time spans (since the early 1970s preferably).

Figure 1: The German Business Cycle



Note: Production index: 5-month centered moving average, Before 1991: West German data spliced in. Shaded areas: Recession phases. Data sources see Table 1.

2. *Vulnerability to revisions:* Several macroeconomic time series are very much prone to revisions (Croushore, 2011). Real-time data sets for Germany, however, are only available for short time periods. We, therefore, restrict our analysis to indicators which are timely observable and less vulnerable to revisions.

3. *Monthly availability:* Because we want to analyze the forecasting performance of leading indicators at different forecasting horizons and, furthermore, because we want to have available as many observations as possible, we restrict our analysis to monthly data.

4. *Granger causality at business-cycle frequency:* König and Wolters (1972); Wolters and Lankes (1989); Wolters et al. (1990); Wolters (1996); Fritsche and Stephan (2002) argue that a necessary condition is that an indicator shows a high coherence at business-cycle frequency with a representative measure of the cycle. A sufficient condition for a reasonable leading indicator is the well known concept of Granger causality (Granger, 1969).

Breitung and Candelon (2006) developed a formal and easy-to-apply test for *Granger causality* in a frequency-domain setting. Application of their test to all indicators with respect to the annual change in industrial production results in a rejection of the null hypothesis of no causality for all indicators at least at some business-cycle frequencies, with the U.S. short-term interest rate being an exception. Detailed test results along with a figure showing the leading indicators and a table summarizing the results of unit-root tests are reported at the end of the paper (Appendix).

The *unit-root tests* show that the transformations of the leading indicators detailed in Table 1 result in stationary time series. One exception is the inflation rate, which may be nonstationary. Estimating the BRT approach on the first difference of the inflation rate gives results (not reported, but available upon

Table 1: Data & Sources

Series	Source	Link	Transformation	Group	ID
Recession phases	Economic Cycle Research Institute	Link	None	None	RECESSION
Short term interest rates	Deutsche Bundesbank	Link	None	Financial	RK
Interest rate spread	Deutsche Bundesbank	Link	None	Financial	SPREAD
Order inflow industry	Deutsche Bundesbank	Link	yoy	Real	ORDER
ifo business climate	FRED / ifo institute	Link	None	Survey	CLIMATE
ifo business climate, current situation	FRED / ifo institute	Link	None	Survey	CLIMATE_SIT
ifo business climate, expectations	FRED / ifo institute	Link	None	Survey	CLIMATE_EXP
Consumer price index (CPI)	Deutsche Bundesbank	Link	yoy	Prices	INF
OECD Stock Market Index	OECD Monthly Economic Indicators	Link	yoy	Financial	STOCK
Industrial production	Deutsche Bundesbank	Link	yoy	Real	PRODUCTION
OECD Consumer Confidence	OECD Monthly Economic Indicators	Link	None	Survey	CONSUMER_CONF
U.S. Industrial production	OECD Monthly Economic Indicators	Link	yoy	Real	US_PROD
Crude Oil Prices: West Texas Intermediate (WTI)	FRED	Link	yoy	Prices	OIL
OECD leading indicator	OECD Monthly Economic Indicators	Link	None	Composite	OECD_LEADING
Real Narrow Effective Exchange Rate for Germany	FRED	Link	yoy	Prices	REER
U.S. Effective Federal Funds Rate	FRED	Link	None	Financial	RK_US

Note: yoy denotes change over previous year. FRED: Federal Reserve Bank of St. Louis database.

request) that are qualitatively similar to the results we shall report in Section 4. The unit-root tests use maximum number of monthly data available because it is well-known that the power of unit-root tests depends on the length of the time period covered by the data rather than on the number of observations given a time period (see Sephton, 1995). Because data availability differs across leading indicators, we also report the start period for every leading indicator. While data for some leading indicators trace back to 1960, data on the real effective exchange rate are available starting in 1964 and data on consumer confidence are available starting in 1973. Because the real effective exchange rate could be an important leading indicator for an open economy like Germany and consumer confidence is a popular leading indicator, we use in our empirical analysis mainly data starting in 1973/1, but in Section 4.9 we shall estimate a smaller model that does not feature the real effective exchange rate and consumer confidence on an extended sample period starting in the early 1960s.

Table 2: Results for Individual Probit Models

Leading indicator	Pseudo R^2	Pseudo R^2	AUROC	S.E.	AUROC	S.E.
Forecast horizon	3 months	6 months	3 months		6 months	
RK	0.3578	0.3943	0.8527	0.0214	0.8668	0.0205
SPREAD	0.3735	0.4247	0.8773	0.0198	0.8944	0.0186
ORDER	0.0648	0.0193	0.7453	0.0262	0.6468	0.0284
CLIMATE	0.1271	0.0447	0.7449	0.0262	0.6438	0.0284
CLIMATE_EXP	0.2021	0.0955	0.8278	0.0228	0.7347	0.0265
CLIMATE_SIT	0.0990	0.0309	0.7162	0.0270	0.6230	0.0287
INF	0.2403	0.2424	0.8086	0.0237	0.8108	0.0236
STOCK	0.1856	0.1269	0.8267	0.0229	0.7871	0.0247
PRODUCTION	0.0415	0.0117	0.6903	0.0276	0.5992	0.0289
CONSUMER_CONF	0.1117	0.0411	0.7350	0.0265	0.6438	0.0284
US_PROD	0.0771	0.0311	0.7525	0.0260	0.6970	0.0275
OIL	0.0271	0.0484	0.5464	0.0290	0.5908	0.0289
OECD_LEADING	0.1348	0.0452	0.7896	0.0246	0.6751	0.0279
REER	0.0015	0.0012	0.5456	0.0290	0.5460	0.0290
RK_US	0.0510	0.0696	0.5992	0.0289	0.6349	0.0285

Note: The Pseudo R^2 is defined as (McFadden, 1974): Pseudo $R^2 = 1 - LL/LL_0$, where LL denotes the value of the maximized log likelihood function of the estimated Probit model, and LL_0 denotes the value of the maximized log likelihood function of a Probit model that features only a constant.

In Table 2, we present a baseline analysis of the *predictive power* of individual leading indicators as measured in terms of the Pseudo R^2 and AUROC statistics. Both statistics are derived using univariate Probit models estimated on the full sample of data and assuming a fixed forecast horizon of three and six months, where the models feature an intercept and a leading indicator as the only variables. The AUROC statistics are computed for the estimated recession probabilities.

The short-term interest rate and the term spread have the largest Pseudo R^2 and AUROC statistics, followed by stock market returns and the inflation rate. The business-climate indicators Pseudo R^2 and AUROC statistics imply that the indicators are mainly useful at a forecast horizon of three months. The oil price, the real effective exchange rate, and the short-term U.S. interest rate have the lowest AUROC statistics at both forecast horizons, and their

Pseudo R^2 indicates a rather poor fit. Hence, while the majority of leading indicators have at least some power for predicting recessions, the informational content differs across leading indicators and there is also some variation across forecast horizons. In Section 4, we shall analyze this variation across leading indicators and forecast horizons in more detail using the BRT approach, and we shall analyze how the in-sample results reported in Table 2 change when we turn to the analysis of out-of-sample forecasts. Moreover, because combining leading indicators may have the potential to increase model fit we shall use the BRT approach to study how the leading indicators can be combined in a unified forecasting model in a meaningful way and how such a unified forecasting model can take into account potential interaction effects of the leading indicators.

4 Empirical Results

4.1 Model Calibration

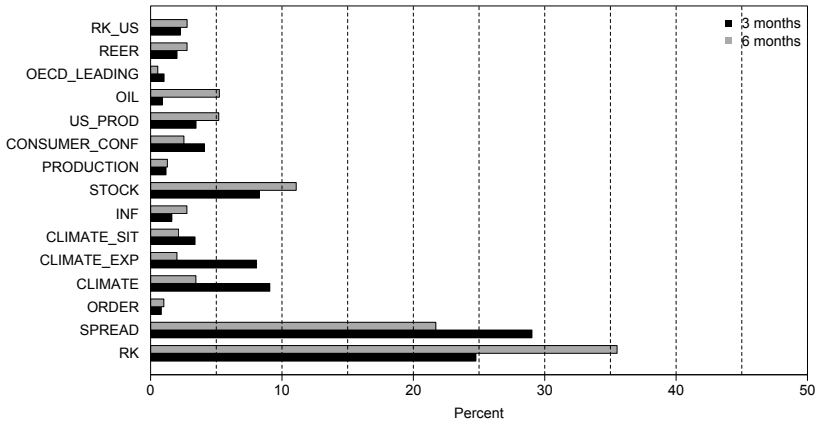
We use the R programming environment for statistical computing (R Core Team, 2015) to carry out our empirical analysis. For estimation of the BRT model, we use the add-on package "gbm" (loss function "Adaboost" Ridgeway, 2015). We allow for a maximum tree depth of 5, which should suffice to capture potential interaction effects. The minimum number of observations per terminal node is 5 and the shrinkage parameter assumes the value $\lambda = 0.005$, where reasonable larger values give similar results. Our estimation strategy is to use 70% of the data (sampling without replacement) to train the BRT model and 30% for quasi out-of-sample model testing. We simulate this process 1,000 times to make statistical inference, where we use five-fold cross validation to determine the optimal number of weak learners in every simulation run. In this respect, we fix the maximum number of weak learners at $M = 3,000$, but the cross-validated bias-variance minimizing number of weak learners is typically much smaller. Across all 1,000 simulations, the average number of optimal weak learners is approximately 634 (standard deviation of roughly 118) when we study a forecasting horizon of three months, and roughly 688 (standard deviation of about 121) when we study a forecasting horizon of six months. Finally, we implement the stochastic-gradient-descent boosting algorithm (Friedman, 2002) by selecting 50% of the training data at random in Step 3a of the gradient-descent-boosting algorithm to built the next weak learner in the recursion.

4.2 Relative Importance of Leading Indicators

The relative importance of a leading indicator in a regression tree is defined as the sum over nonterminal nodes of the squared improvement resulting from using a leading indicator to form splits (Breiman et al., 1984), where this definition extends to boosted tree ensembles by averaging across weak learners (Friedman, 2001). In our application, we further average across simulation runs.

Figure 2 shows that, according to this definition, the short-term interest rate and the term spread are the two most influential leading indicators. While the term spread is relatively more important if the forecasting horizon is three months, the short-term interest rate is relatively more important if the forecast horizon is six months. The relative importance of stock market returns and

Figure 2: Relative Importance of Leading Indicators



business-climate indicators ranges between 8% to 10% for a forecast horizon of three months. While the relative importance of the business-climate indicators decreases when we switch to a forecast horizon of six months, the relative importance of stock market returns increases somewhat to approximately 11%. For U.S. data, Estrella and Mishkin (1998) also find that stock prices have predictive value for recessions, especially at forecast horizons from one quarter to three quarters, and Farmer (2012) emphasizes the predictive value of the stock market for the unemployment rate. The other leading indicators have a relative importance of at or below roughly 5%. Hence, although they are much less important than the short-term interest rate and the term spread, the BRT approach occasionally uses the informational content of these other leading indicators for tree building.

The general message of the relative importance plot reported in Figure 2 does not change when we analzye a dynamic model by adding three lags of every variable to our vector, $\mathbf{x}_t$, of leading indicators (results not reported, but available upon request). This robustness stems from the insensitivity of the BRT approach to the inclusion of irrelevant predictors in the set of leading indicators. Its robustness is an important feature of the BRT approach given that earlier researchers have found that "overfitting is a serious problem in macroeconomic predictions. Even when only a few variables are used, the addition of a single variable or another lag of a variable can undermine the predictive power of a parsimonious model" (Estrella and Mishkin, 1998, page 55).

We also note that we assume throughout our analysis that, as in the research by Ng (2014), our boosted forecasting models do not feature lags of the response variable in the set of leading indicators. Modifying the BRT approach to incorporate lags of the response variables in the set of predictors requires few modifications relative to our analysis because the BRT approach can handle different types of predictors. At the same time, adding lags of the response variable complicates the analysis because forecasts computed by means of such a modified BRT approach should be compared with forecasts from a dynamic

Probit model. There are alternative ways of how to incorporate dynamics into Probit models (lagged recession indicator, lagged latent state of the economy, interaction terms; see Kauppi and Saikkonen, 2008). Moreover, computing forecasts becomes more difficult for a dynamic model and, in addition, the question of when exactly a recession indicator is known to a forecaster such that he or she can use the indicator to compute forecasts needs to be carefully addressed (see Nyberg, 2010, who assumes a lag of nine months).

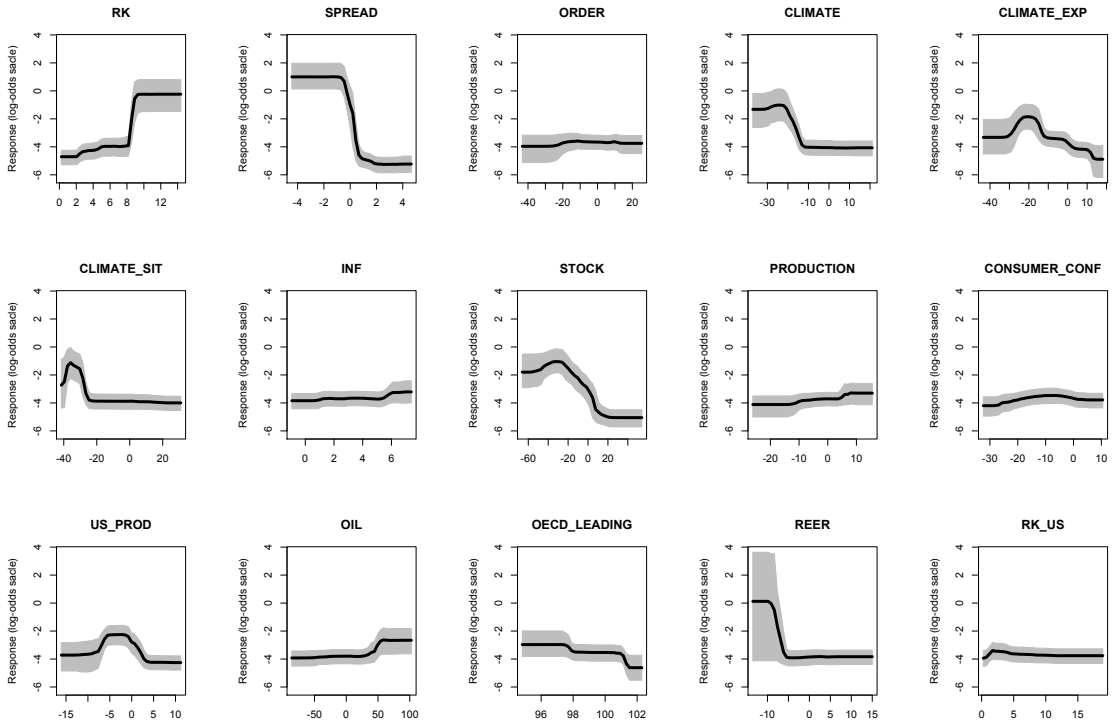
4.3 Marginal Effects

Figures 3 and 4 (for forecast horizons of three and six months) show marginal effects obtained by fitting the BRT model on the training data, where the shaded areas denote 95% confidence intervals computed across 1,000 simulation runs. The marginal effects show the effect of a leading indicator (horizontal axis) on the probability of a recession (vertical axis; measured at log-odds scale), where the effects of the other leading indicators are controlled for using the weighted-traversal technique described by Friedman (2001, page 1221). At the end of the paper (Appendix), we plot histograms which shed some light on how often the data visited different parts of the marginal-effect curves historically (although their is no simple one-to-one correspondence because of the hierarchical structure of the weak learners; see Section 4.4 on interaction effects).

While the recession probability only slightly increases for low short-term interest rates, it shows an abrupt increase, depending on the forecasting horizon, at around 6%−8% and stays thereafter constant for higher interest rates. A negative term spread is associated with a higher recession probability than a positive term spread, where the log-odds ratio switches from a positive to a negative value when the term structure is flat. The marginal effects further show with respect to the stock market that the recession probability is lower in times of bull markets and substantially increases in times of bear markets. The marginal effects also recover how the recession probability gradually decreases in the range of stock market returns from −40% to 20%.

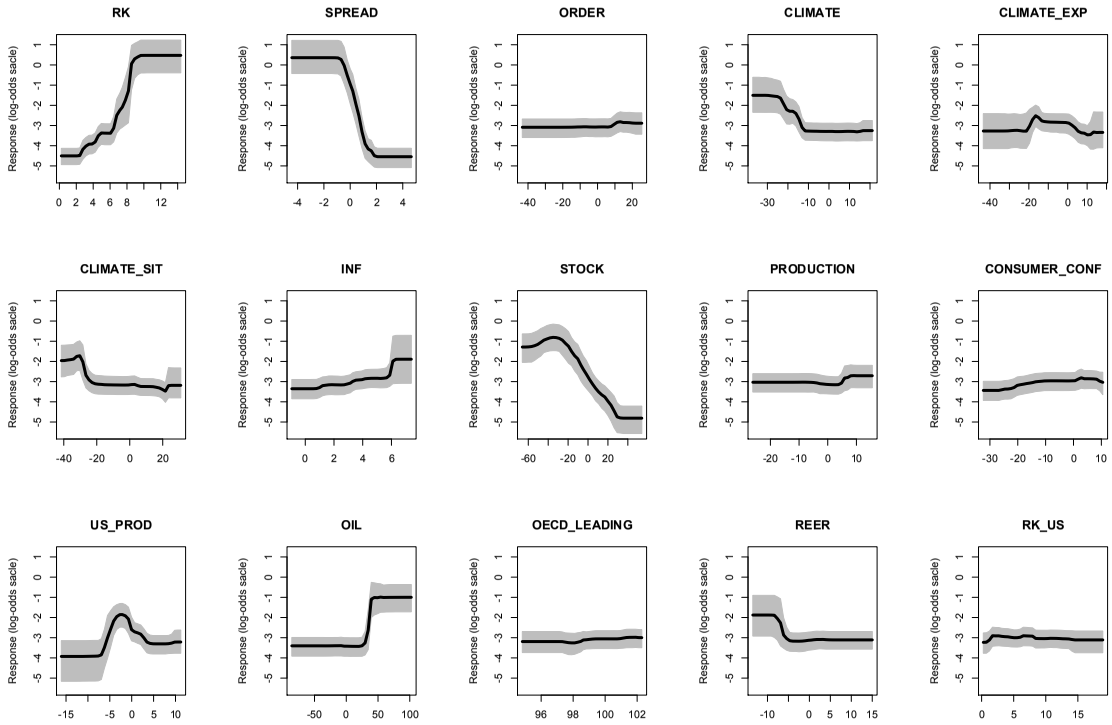
The marginal effect estimated for order inflows is more or less flat, while larger values of the business-climate indicators lead to a lower recession probability, an effect that is somewhat more visible for a forecasting horizon of three months. The recession probability is rather insensitive to variations in the inflation rate (forecasting horizon three months), production growth, the consumer-confidence indicator, the OECD leading indicator (especially at a forecasting horizon of six months), and the short-term U.S. interest rate. The recession probability increases when the oil price increases, and this effect is stronger at a forecasting horizon of six months. Fluctuations of the real effective exchange rate hardly affect the recession probability, where the marginal effect for depreciations is estimated with less precision than the marginal effect for appreciations.

Figure 3: Marginal Effects (Forecast Horizon: 3 Months)



Note: The horizontal axis shows the mean per quantile of 2.5% width of the leading indicators computed across all simulation runs. The thick line shows the corresponding mean per quantile of the log-odds ratio computed across all simulation runs. The gray area denotes the 95% confidence interval computed for every quantile.

Figure 4: Marginal Effects (Forecast Horizon: 6 Months)



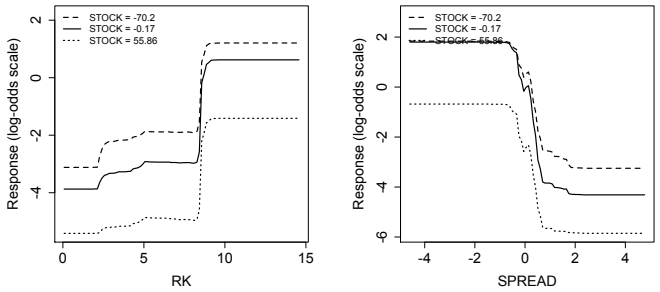
Note: The horizontal axis shows the mean per quantile of 2.5% width of the leading indicators computed across all simulation runs. The thick line shows the corresponding mean per quantile of the log-odds ratio computed across all simulation runs. The gray area denotes the 95% confidence interval computed for every quantile.

4.4 Interaction Effects

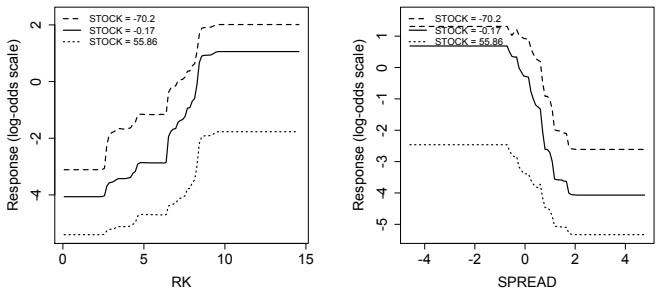
Figure 5 shows the recession probability (log-odds scale) as a function of the short-term interest rate and the term spread for alternative realizations of stock-market returns, where the BRT model is estimated on the full sample of data. Stock market returns assume their sample minimum (bear market), their sample mean (neutral market), and their sample maximum (bull market). Figure 5 shows that the recession probability corresponding to a high short-term interest rate is smaller (and, in fact, the log-odds ratio is negative, implying that the recession probability is smaller than 0.5) in times of a bull market than in times of a bear market. Similarly, a positive term spread is associated with a smaller recession probability in times of a bull market. Results further show that the difference between the marginal effects computed for a bear market and a neutral market is smaller than the difference between the marginal effects computed for a neutral market and a bull market, highlighting the dependence of the response of the recession probability on stock market conditions.

Figure 5: Interaction Effects: Influence of the Stock Market

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months



4.5 ROC Analysis

Table 3 informs about the sampling distribution of the AUROC statistic as computed from the simulated 1,000 samples of the quasi out-of-sample experiment (30% test data). Apparently, the BRT approach yields a very good fit. The AUROC statistic assumes values above 0.9 in all simulations, implying ROC

Table 3: AUROC Statistics for the BRT Approach

Forecast horizon	Mean	CI lower bound	CI upper bound
BRT Approach			
3 months	0.9891	0.9657	0.9989
6 months	0.9823	0.9600	0.9958
BRT versus Simple Probit			
3 months	0.1307	0.0692	0.2030
6 months	0.0959	0.0488	0.1643
BRT versus BMA Probit			
3 months	0.1227	0.0624	0.1937
6 months	0.0912	0.0449	0.14865

Note: CI = 95% confidence interval. Test fraction: 30%. Number of simulation runs: 1,000. For the comparisons of the BRT approach with the Probit approach, mean denotes the average difference between the AUROC statistic for the BRT approach minus the AUROC statistic for the Probit approach.

curves that settle in the vicinity of the north-western corner of a unit quadrant.

In order to put our results into perspective, we subject a simple Probit model featuring all leading indicators to the same quasi out-of-sample forecasting experiment, where we use the same random seed of training and test data. We then subtract the simulated AUROC statistics for the Probit approach from the AUROC statistics computed for the BRT approach. Table 3 shows that the difference of the AUROC statistics is significantly positive.

Comparing the BRT approach with a simple Probit approach is a bit unfair. The BRT approach combines the leading indicators by means of an ensemble of regression trees to form recession forecasts, where the process of growing trees accounts for the predictive power of the leading indicators. The simple Probit approach, in contrast, throws all leading indicators into a single model without weighting the leading indicators according to their importance. We, therefore, next compare the BRT approach with a more sophisticated Probit approach. To this end, we estimate on the training data obtained from the simulation runs of our quasi out-of-sample forecasting experiment for every leading indicator a Probit model. We then use the estimated Probit models to forecast recessions using the test data generated in the context of our quasi out-of-sample forecasting experiment. Finally, we use Bayesian model averaging to build a single recession forecast by combining weighted forecasts from the individual Probit models. Specifically, like Berge (2015), we assume that the a priori probability of each model is the same, and use a result derived by Raftery (1995, page 145) to approximate the a posteriori probability, post_i, of model i by means of the Bayesian Information Criterion (BIC, Schwarz 1978) as follows:

$$\text{post}_i = \exp(-\frac{1}{2}\text{BIC}_i) / \sum_{i=1}^{N(\mathbf{x}_t)} \exp(-\frac{1}{2}\text{BIC}_i), \tag{16}$$

where $N(\mathbf{x}_t)$ denotes the number of leading indicators in $\mathbf{x}_t$. The BIC is computed as $\text{BIC}_i = -2LL_i + 2n$, where LL_i denotes the maximized log likelihood function of Probit model i, and the product $2n$ denotes the number of parameters (in our case, an intercept plus the coefficient of the leading indicator being studied under model i) times the number of observations in the training data.

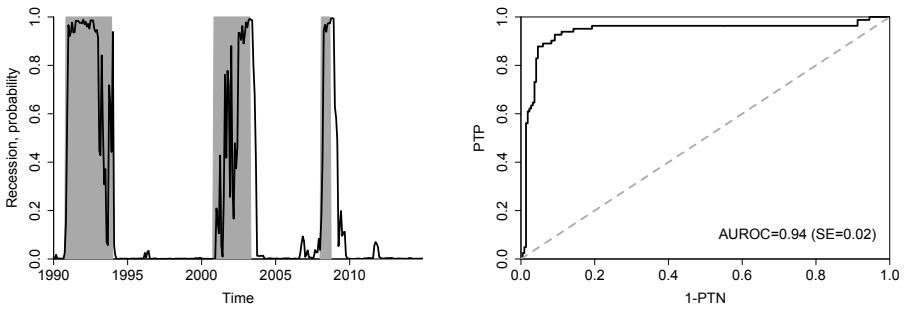
Table 3 shows the mean of the sampling distribution of the difference between the AUROC statistic for the BRT approach minus AUROC statistic for the Bayesian model averaging (BMA) Probit approach along with the 95% confidence interval of the sampling distribution. Simulation results (Table 3) show that the BRT approach outperforms the BMA Probit approach in terms of the AUROC statistic.

4.6 Recursive Estimation

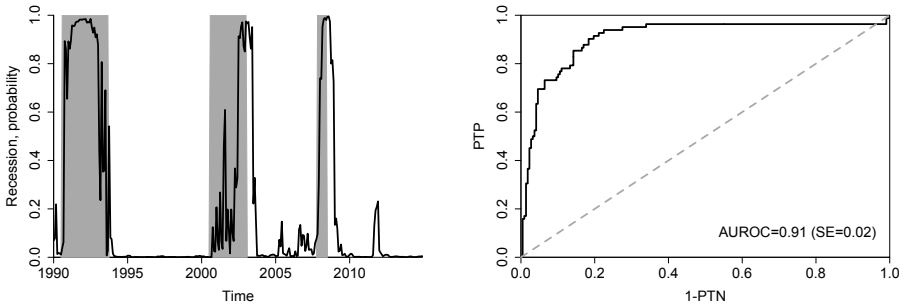
The quasi out-of-sample forecasting experiment requires drawing random samples from the full sample of data to train the model. The remaining test data are scattered across the sample so that information from business-cycle developments that occurred later in the sample are used in a kind of backtesting experiment to study how the model performs on the test data. Hence, the quasi-out-of-sample forecasting experiment most likely overestimates the out-of-sample performance of the BRT approach (and the Probit approach). We, therefore, present in Figure 6 results of an alternative, and empirically plausible, out-of-sample experiment. We start this experiment by estimating the BRT model on data up to and including 1989/12. We then use data for the next quarter to make three monthly forecasts with the estimated BRT model. At the end of the quarter, we reestimate the BRT model and use the reestimated model to make the next three forecasts. We continue this recursive forecasting and updating process until we reach the end of the sample period.

Figure 6: Out-of-Sample Performance

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months

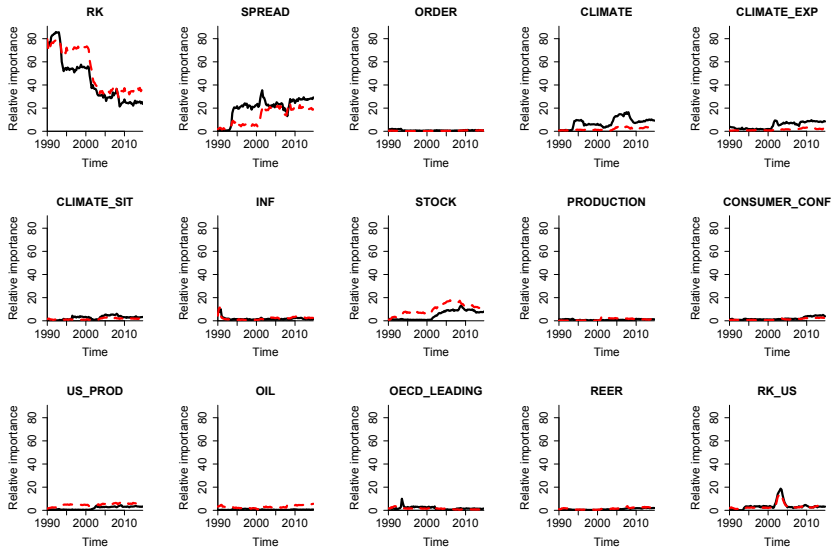


The results show that the model captures the beginning of the recession of the early 1990s well. It also describes accurately the end of the recession, where the recession probability gets unstable towards the end of the recession. It is also interesting to note that the model accurately signals the beginning and the end of the *Great Recession*. At the start of the recession of the early 2000s, the recession probability needs some time to gradually build up. The recession probability also shows some remarkable upticks and downticks during the recession. Notwithstanding, the AUROC statistic assumes a relatively large value of 0.94 (0.91) for a forecasting horizon of three (six) months, where the 95% confidence interval ranges from 0.90 to 0.98 (0.86 to 0.95). As expected, the overall out-of-sample forecasting performance of the BRT approach, thus, is somewhat less impressive than in the backtesting quasi out-of-sample forecasting experiment. The out-of-sample forecasting performance of the BRT approach, however, still dominates a pure noise signal.

4.7 Variation Across Time

Figure 7 shows how the relative importance of the leading indicators has changed since 1989/12 (on the stability of prediction models for U.S. and German inflation and real activitiy, see Estrella et al., 2003). The results for a forecast horizon of three months show a trend decline of the relative importance of the short-term interest rate from over 80% in the early 1990s to slightly above 20% (see Figure 2) when the end of the sample period is reached. Results for a forecast horizon of six months are similar, but the relative importance of the short-term interest rate only drops to approximately 35%. The relative importance of the term spread, in contrast, substantially increases from a negligibly small importance in the early 1990s to approximately 30% when the forecast horizon is three months and about 20% when the forecast horizon is six months.

Figure 7: Changing Relative Importance of Leading Indicators

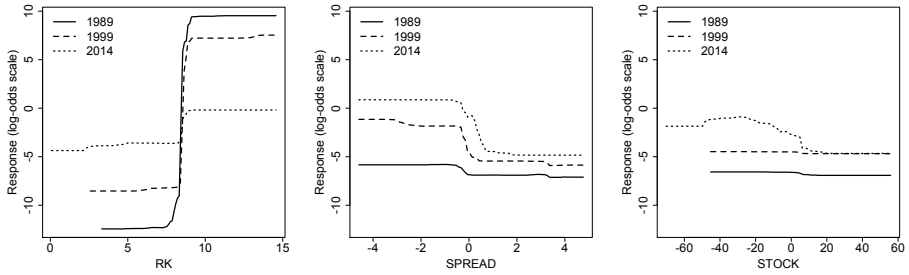


Note: Black (red) solid (dashed) lines: Forecast horizon 3 (6) months.

While the relative importance of the other leading indicators has stayed more or less constant at a low level over the years, the relative importance of the business climate indicators slightly increased, but only for a forecasting horizon of three months. We also observe that the relative importance of U.S. production growth has increased somewhat over time, an effect that is stronger for a forecasting horizon of six months than for a forecasting horizon of three months. The relative importance of stock market returns has experienced a more substantial increase at both forecasting horizons. Consistent with the results plotted in of Figure 2, the relative importance of stock market returns reaches about 8% at the end of the sample for a forecast horizon of three months, and approximately 11% when the forecasting horizon is six months.

Figure 8: Changing Marginal-Effect Curves

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months

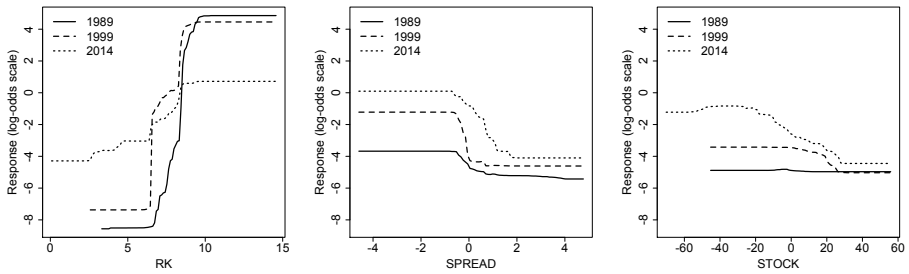


Figure 8 shows how the marginal effects for the short-term interest rate, the term spread, and stock market returns have changed over time. The marginal effects plotted in the figure are based on estimates of the BRT model on the full sample period (including data up to and including 2014/12) and two shorter sample periods (including data up to and including 1989/12, 1999/12). The function that summarizes the marginal effects of the short-term interest rate undergones a clockwise rotation as the sample period becomes longer. The marginal-effect curve for the term spread, in contrast, shifts upward over time, where the shift is more pronounced for a negative term spread than for a positive term spread. The marginal-effect curve for stock market returns also exhibits a shift across time. The recession probability in the longer sample periods is higher in times of a bull market than in the shorter sample period (1989) for a forecast horizon of three months. In addition, the recession probability in times of a bear market is larger, in case of both forecast horizons, when we estimate

the model on data ending in 2014/12 than in the two shorter sample periods.

4.8 Optimal Cutoff Value

We now use the ROC curves implied by the BRT approach to select an optimal cutoff value, c. We start with an additive utility function given by (see Baker and Kramer, 2007; Berge and Jordà, 2011)

$$U(c) = U_{11}\pi PTP(c) + U_{01}\pi(1 - PTN(c)) \\ + U_{10}(1-\pi)(1-PTN(c)) + U_{00}(1-\pi)PTN(c), \quad (17)$$

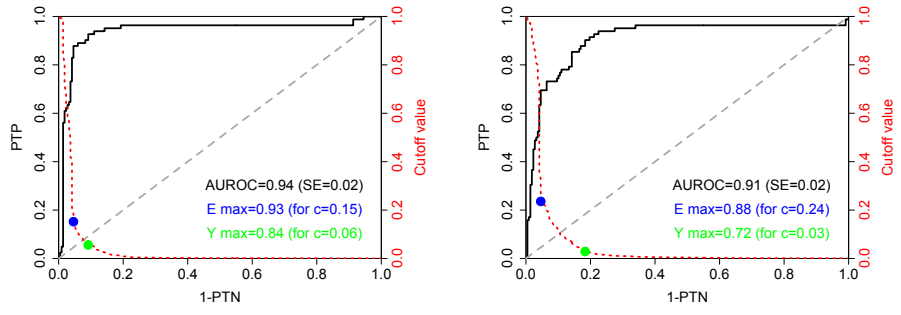
where π denotes the unconditional probability of a recession and U_{ij} denotes a forecaster's utility if the forecast is i and the economy is in state j. The first-order condition for a maximum yields

$$\frac{\partial PTP(c)}{\partial r} = \text{utility ratio} \times \frac{1-\pi}{\pi}, \quad (18)$$

where $r = 1 - PTN(c)$ and utility ratio $= (U_{00} - U_{10})/(U_{11} - U_{01})$. Equation (18) stipulates that the optimal cutoff value can be derived by equating the slope of the ROC curve to the ratio of net utility of forecasts in non-recession periods and the net utility of forecast in recession periods (see also Berge and Jordà, 2011).

If we assume that a forecaster chooses a cutoff value so as to maximize a utility function that is symmetric in the sense such that utility ratio $= 1$, then maximizing utility is equivalent to maximizing the efficiency, E of forecasts defined as $E(c) = \pi PTP(c) + (1-\pi)PTN(c)$ (for a survey of approaches studied in the ROC literature to select a cutoff value, see Greiner et al., 2000). If we invoke the further assumption that the unconditional probability of a recession is $\pi = 0.5$, then the utility-maximizing cutoff value maximizes the Youden index (Youden, 1950), which is defined as $Y(c) = PTP(c) + PTN(c) - 1$. For our data, the Youden index is suboptimal because we have $\pi = 0.28$.

Figure 9: Optimal Cutoff Values



Note: Results for the recursive out-of-sample experiment. E = efficiency. Y= Youden index. Left panel: forecast horizon 3 months. Right panel: forecast horizon 6 months.

Figure 9 shows the cutoff values that maximize the efficiency of forecasts and the Youden index for the recursive out-of-sample experiment. Maximizing

Table 4: Optimal Cutoff Values in the Quasi out-of-Sample Experiment

Index / Forecast horizon	3 months	6 months
Efficiency	0.9682	0.9514
Cutoff value	0.3604	0.3303
Youden index	0.9351	0.8964
Cutoff value	0.2316	0.1740

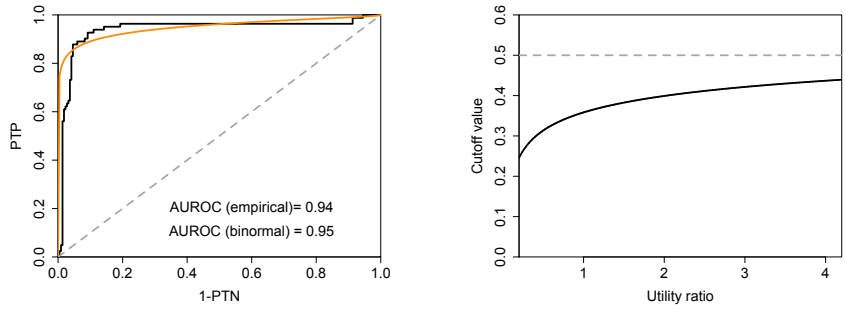
Note: Means computed across 1,000 simulation runs. Test fraction: 30%.

the effeciency of forecasts requires that a forecaster sets a cutoff value of 0.15 (0.24) for a forecast horizon of three (six) months. Hence, when the forecast horizon is fixed at three months, the decision rule is: "Forecast a recession when the recession probability implied by the BRT approach reaches 0.15." For the Youden index, the decision rule is to predict a recession whenever the recession probability implied by the BRT model reaches the cutoff value 0.06 (0.03) when the forecast horizon is three (six) months. This cutoff value is implausibly low, but it reflects the curvature of the steep ROC curves shown in Figure 9.

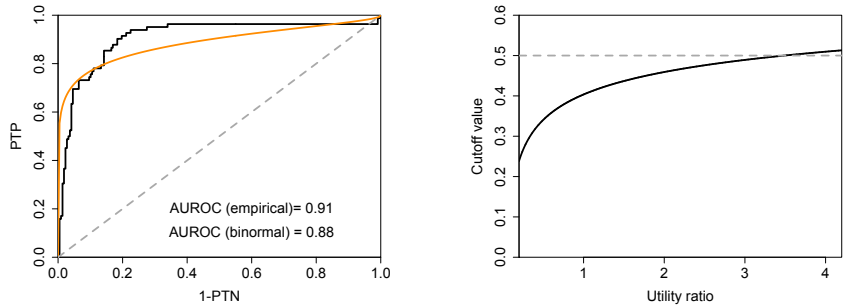
Table 4 summarizes the implications of our quasi out-of-sample experiment for the optimal cutoff values. The average optimal cutoff value that maximizes the efficiency of forecasts is 0.35 and 0.22 for the two forecast horizons, and 0.34 and 0.18 for the Youden index. Further inspection showed that the optimal cutoff values display substantial variability across simulation runs.

Figure 10: Asymmetric Loss and the Optimal Cutoff Value

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months



Because studying business-cycle forecasts under asymmetric loss has become increasingly popular in recent research (Elliott et al., 2008; Döpke and Fritsche, 2010; Pierdzioch et al., 2015), we relax in Figure 10 the assumption that the utility function is symmetric in the sense described above. To this end, we use a binormal model to approximate the slope of the discrete ROC curve depicted in Figure 9 and then choose the optimal cutoff value to satisfy Equation (18) for alternative values of the utility ratio (for a survey of various approximation methods and references to the relevant literature, see Marzban, 2004). If the utility of true recession forecasts is the same as the utility derived from true non-recession forecasts, then utility ratio < 1 indicates that the disutility from missing a recession exceeds the disutility from wrongly predicting a recession when none occurs. Hence, if the disutility of missing a recession is relatively large the optimal cutoff value is small, while the optimal cutoff value is large if the disutility of predicting a recession that does not happen is relatively large.

The binormal model fits the ROC curve better for a forecast horizon of three months than for a forecast horizon of six months. For both forecast horizons the optimal cutoff value is an increasing function of the utility ratio (Figure 10). For a forecast horizon of six months, the optimal cutoff value even crosses the 0.5 line and eventually becomes positive when the utility ratio becomes large. Hence, a cutoff value of 0.5 can be optimal under asymmetric loss when a symmetric loss function requires a deviation from such a cutoff value.

4.9 Robustness Checks

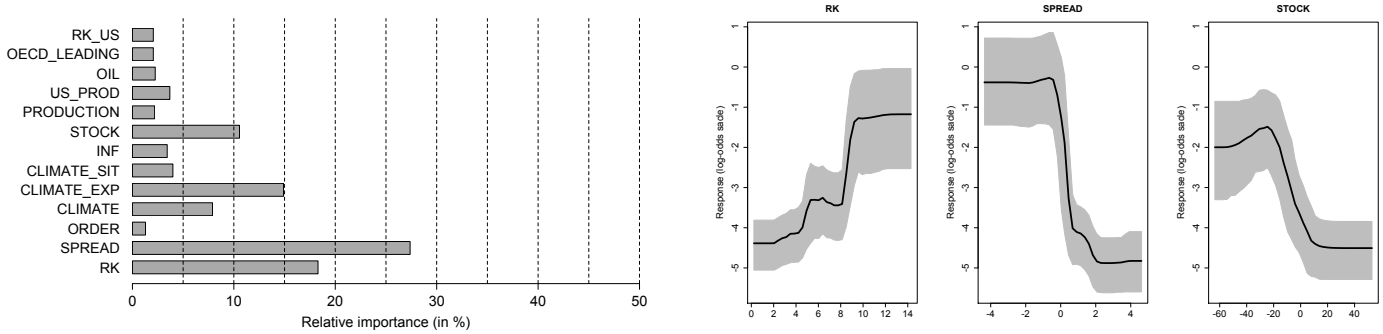
Apart from data on the real effective exchange rate and consumer confidence, data for the other leading indicators trace back to the early 1960s (Section 3.2). Next, we use this additional data for a robustness check. First, we use our result that the real effective exchange rate and consumer confidence are not among the top leading indicators. We delete the real effective exchange rate and consumer confidence from our list of leading indicators and extend our sample period, which now starts in 1961/1. Second, we use the modified dataset to change the calibration of the BRT model. Specifically, we reduce the proportion of random training data to 50%, implying that the quasi out-of-sample forecasting scenario is now more challenging than in the previous subsections.

Figure 11 shows that the short-term interest rate and the term spread are the top leading indicators. For a forecast horizon of six months, the interest-rate indicators are followed by the stock market. For a forecast horizon of three months, the stock market competes with the business climate indicators, and the dominance of the interest-rate indicators further strengthens. Figure 11 further shows that adding more than ten years of monthly data to our sample period has left the marginal-effects curves for the interest-rate indicators and stock market returns qualitatively unaffected.

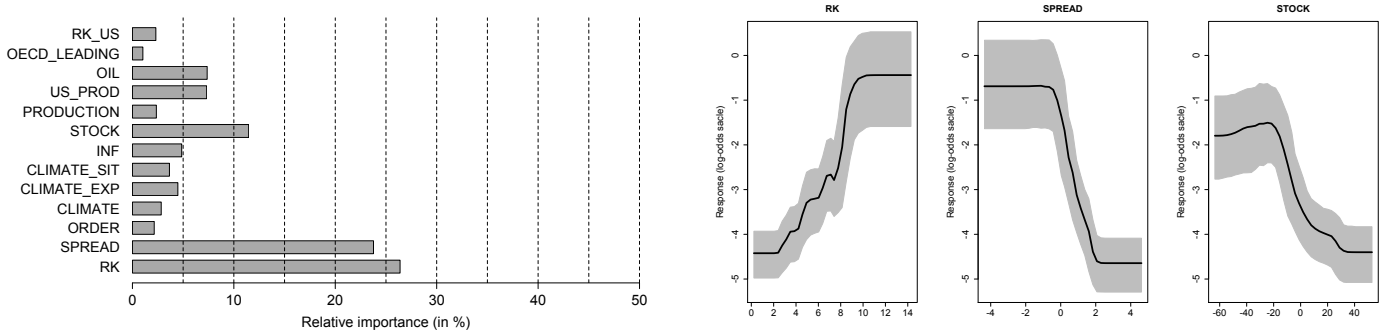
Table 5 shows that the competitive performance of the BRT approach relative to a simple and a BMA Probit approach also can be observed for the extended dataset and the modified model.

Figure 11: Marginal Effects for the Modified Model

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months



Note: Results for the modified model. The random training data comprises 50% of the data. The extended sample period starts in 1961/1. Number of simulation runs: 1,000.

Table 5: Comparison of the BRT Approach With a Probit Approach (AUROC, Modified Model)

Forecast horizon	Mean	CI lower bound	CI upper bound
BRT Results			
3 months	0.9721	0.9525	0.988
6 months	0.9635	0.9358	0.981
BRT versus Simple Probit			
3 months	0.1391	0.1011	0.1779
6 months	0.1092	0.0761	0.1481
BRT versus BMA Probit			
3 months	0.1256	0.0681	0.1763
6 months	0.0987	0.0592	0.1446

Note: Results for the modified model. The random training data comprises 50% of the data. The extended sample period starts in 1961/1. Number of simulation runs: 1,000.

Table 6: Results for the QPS Measure (Modified Model)

Forecast horizon	Mean	CI lower bound	CI upper bound
BRT Results			
3 months	0.0594	0.0388	0.0844
6 months	0.0701	0.0456	0.0989
BRT versus Simple Probit			
3 months	-0.0626	-0.085	-0.0436
6 months	-0.0470	-0.064	-0.0290
BRT versus BMA Probit			
3 months	-0.0599	-0.0793	-0.0414
6 months	-0.0449	-0.0636	-0.0262

Note: Results for the modified model. The random training data comprises 50% of the data. The extended sample period starts in 1961/1. Number of simulation runs: 1,000.

As another robustness check, we use the quadratic probability score (QPS) as an alternative performance measure to compare the BRT approach with the Probit approach. The QPS is defined as

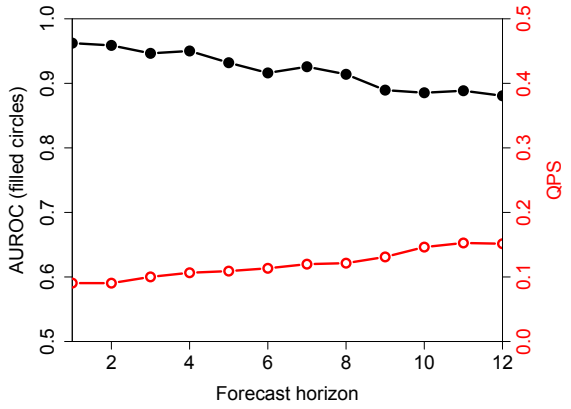
$$QPS = \frac{1}{N}\sum_{i=1}^{N}(\hat{P}(y_{t+k} = 1|\mathbf{x_t}) - \mathbf{1}_{y_{t+k}=1})^2, \tag{19}$$

which can assume values between 0 and 1, where smaller values signal better forecasting performance (for applications, see Kaminsky, 2006; Lahiri and Wang, 2013; Savona and Vezzoli, 2015). Upon using our extended sample period, we compute the QPS for the BRT approach and then subtract the QPS for the Probit approach, where we use our quasi out-of-sample experiment to compute the sampling distribution of the resulting relative QPS measure. A negative relative QPS measure indicates that the BRT approach dominates the Probit approach.

The results summarized in Table 6 show that the relative QPS is negative for both the simple Probit model and the BMA Probit model, and the simulated 95% confidence intervals indicate that the performance of the BRT approach is

significantly better than the performance of the Probit approach.

Figure 12: Results for Alternative Forecast Horizons



Note: Results for the out-of-sample forecasting experiment. First forecast: 1980/1. The modified model is updated every three months.

As a final robustness check, we plot in Figure 12 results for the AUROC statistic and the QPS for alternative forecast horizons of up to one year. The results are based on a recursive out-of-sample forecasting experiment, where we analyze the extended dataset, which starts in 1961/1. The first forecast is for 1980/1 and the model is recursively updated every three months. The results show that the forecasting performance does not substantially deteriorate when we consider a forecasting horizon somewhat beyond the three and six months studied in the other parts of this research. The QPS function moderately increases in the forecasst horizon and the AUROC statistsic remains safely above the pure noise benchmark of 0.5 even for a forecast horizon of twelve months.

5 Concluding Remarks

From the viewpoint of applied business-cycle forecasting, machine-learning techniques are not a substitute for experience in business-cycle forecasting in general and in interpreting changes in estimated recession probabilities in particular. Our empirical results, however, show that machine-learning techniques can yield important insights into how an economy works. The BRT approach that we have studied in this research is a machine-learning technique that yields a competitive forecasting approach as compared to variants of the popular Probit approach. The BRT approach can complement the Probit approach as an instrument useful for practical business-cycle analysis, but it also has additional features. One feature is that the BRT approach is a natural modelling platform for analyzing the relative importance of leading indicators. Another feature is that the BRT approach makes it possible to study complicated marginal effects of leading indicators on the recession probability. The BRT approach also helps to recover the complex ways in which leading indicators interact in predicting recessions.

Our results show that the short-term interest rate and the term spread are two important leading indicators of recessions in Germany. While the relative importance of the short-term interest rate has decreased over time, the relative importance of the term spread has increased. Interestingly, the relative importance of the stock market has increased somewhat over time. It is interesting to investigate the role of the stock market in future research in more detail using either the BRT approach or some other machine-learning technique. Our results further show that the BRT approach can be a useful technique for the analysis of economic policy. The changes in the relative importance of the short-term interest rate as a leading indicator of recessions that we have detected clearly have implications for monetary policy. Also interesting from a monetary-policy perspective is the relatively small marginal effect of a variation in the short-term interest rate when the short-term interest rate is close to its zero-lower bound. Another result that is of interest for monetary-policy analysis, but also for economic model building, concerns the dependence of the effect of a variation of the short-term interest rate on the recession probability on the state of the stock market. This state dependence deserves further attention in future research.

A natural extension of our research is to broaden the list of leading indicators to include, for example, fiscal-policy indicators and additional asset prices. Asset price that deserves special attention in this respeact are house prices. The BRT approach lends itself to shed light on how movements in house prices affect the probability of an upcoming recession. Because many interesting indicators are not available for Germany for a long time span, we leave such an analysis to future research.

References

Artis, M. J. / *Kontolemis,* Z. G. / *Osborn,* D. R. (1997): "Business Cycles for G7 and European Countries", The Journal of Business 70(2), 249–279.

Bai, J. / *Ng,* S. (2009): "Boosting Diffusion Indices", Journal of Applied Econometrics 24(4), 607–629.

Baker, S. G. / *Kramer,* B. S. (2007): "Peirce, Youden, and Receiver Operating Characteristic curves", The American Statistician 61(4), 343–346.

Bamber, D. (1975): "The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph", Journal of Mathematical Psychology, 12(4), 387–415.

Berge, T. J. (2015): "Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle", Journal of Forecasting 34(6), 455–471.

Berge, T. J. / *Jordà,* Ò. (2011): "Evaluating the Classification of Economic Activity into Recessions and Expansions", American Economic Journal: Macroeconomics 3(2), 246–277.

Breiman, L. (1996): "Bagging Predictors", Machine Learning 24(2), 123–140.

– (2001): "Random Forests", Machine Learning 45(1), 5–32.

Breiman, L. / *Friedman,* J. H. / *Olshen,* R. A. / *Ston,* C. J. (1984): "Classification and Regression Ttrees", Wadsworth and Brooks/Cole, Monterey.

Breitung, J. / *Candelon,* B. (2006): "Testing for Short- and Long-run Causality: A Frequency Domain Approach", Journal of Econometrics 132(2), 363–378.

Bry, G. / *Boschan,* C. (1971): Programmed Selection of Cyclical Turning Points, UMI, chap. 2. 7–63.

Buchen, T. / *Wohlrabe,* K. (2011): "Forecasting with Many Predictors: Is Boosting a Viable Alternative?", Economics Letters 113(1), 16–18.

Bühlmann, P. / *Hothorn,* T. (2007): "Boosting Algorithms: Regularization, Prediction and Model Fitting", Statistical Science 22(4), 477–505.

Burns, A. F. / *Mitchell,* W. C. (1946): "Measuring Business Cycles", National Bureau of Economic Research.

Cottrell, A. / *Lucchetti,* R. (2015): GNU Regression, Econometric and Time-series Library, version 1.9.92 (windows), `http://gretl.sourceforge.net/`. Accessed: May 6, 2015.

Croushore, D. (2011): "Frontiers of Real-Time Data Analysis", Journal of Economic Literature 49(1), 72–100.

Drechsel, K. / *Scheufele,* R. (2012): "The Performance of Short-term Forecasts of the German Economy Before and During the 2008/2009 Recession", International Journal of Forecasting 28(2), 428–445.

Duttagupta, R. / *Cashin,* P. (2011): "Anatomy of Banking Crises in Developing and Emerging Market Countries", Journal of International Money and Finance 30(2), 354–376.

Döpke, J. / *Fritsche,* U. (2010): "Evaluating German Business Cycle Forecasts under an Asymmetric Loss Function", Journal of Business Cycle Measurement and Analysis 2010/1, 1– 18.

Economic Cycle Research Institute (2013): Business Cycle Peak and Trough Dates, 1948-2011, `http://www.businesscycle.com/ecri-business-cycles/international-business-cycle-dates-chronologies`. Date of download: 19.09.2013.

Elliott, G. / *Komunjer,* I. / *Timmermann,* A. (2008): "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?", Journal of the European Economic Association 6(1), 122–157.

Elliott, G. / *Rothenberg,* T. J. / *Stock,* J. H. (1996): "Effcient Tests for an Autoregressive Unit Root", Econometrica 64(4), 813–836.

Estrella, A. / *Hardouvelis,* G. A. (1991): "The Term Structure as a Predictor of Real Economic Activity", The Journal of Finance 46(2), 555–576.

Estrella, A. / *Mishkin,* F. S. (1998): "Predicting US Recessions: Financial Variables as Leading Indicators", Review of Economics and Statistics 80(1), 45–61.

Estrella, A. / *Rodriguez,* A. P. / *Schich,* S. (2003): "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States", Review of Economics and Statistics 85(3), 629–644.

Farmer, R. E. (2012): "The Stock Market Crash of 2008 Caused the Great Recession: Theory and Evidence", Journal of Economic Dynamics and Control 36(5), 693–707.

Freund, Y. / *Schapire,* R. E. (1997): "A Decision-theoretic Generalization of On-line Learning and an Application to Boosting", Journal of Computer and System Sciences 55(1), 119–139.

Friedman, J. H. (2001): "Greedy Function Approximation: A Gradient Boosting Machine", The Annals of Statistics 29(5), 1189–1232.

– (2002): "Stochastic Gradient Boosting", Computational Statistics and Data Analysis 38(4), 367–378.

Friedman, J. H. / *Hastie,* T. / *Tibshirani,* R. (2000): "Additive Logistic Regression: A Statistical View of Boosting", The Annals of Statistics 28(2), 337–407.

Fritsche, U. (1999): "Vorlaufeigenschaften von Ifo-Indikatoren für Westdeutschland", Discussion Papers of DIW Berlin No. 179, DIW Berlin, German Institute for Economic Research, `http://ideas.repec.org/p/diw/diwwpp/dp179.html`.

Fritsche, U. / *Kuzin,* V. (2005): "Prediction of Business Cycle Turning Points in Germany", Jahrbücher für Nationalökonomie und Statistik 225(1), 22–43.

Fritsche, U. / *Stephan,* S. (2002): "Leading Indicators of German Business Cycles", Jahrbücher für Nationalökonomie und Statistik 222(3), 289–315.

Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-spectral methods", Econometrica 37(3), 424–438.

Greene, W. H. (2003): "Econometric Analysis", Prentice Hall.

Greiner, D. / *Pfeiffer,* D. / *Smith,* R. (2000): "Principles and Practical Application of the Receiver-operating Characteristic Analysis for Diagnostic tests", Preventive Veterinary Medicine 45(1), 23–41.

Hanley, J. A. / *McNeil,* B. J. (1982): "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve", Radiology 143(1), 29–36.

Harding, D. / *Pagan,* A. (2003): "A Comparison of two Business Cycle Dating Methods", Journal of Economic Dynamics and Control 27(9), 1681–1690.

Hastie, T. / *Tibshirani,* R. / *Friedman,* J. (2009): "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, Berlin.

James, G. / *Witten,* D. / *Hastie,* T. / *Tibshirani,* R. (2013): "An Introduction to Statistical Learning - with Applications in R", Springer, Berlin.

Kaminsky, G. L. (2006): "Currency Crises: Are They All the Same?", Journal of International Money and Finance 25(3), 503–527.

Kauppi, H. / *Saikkonen,* P. (2008): "Predicting U.S. recessions with dynamic binary response models", Review of Economics and Statistics 90, 777 – 791.

Kholodilin, K. A. (2005): "Forecasting the German Cyclical Turning Points: Dynamic Bi-Factor Model with Markov Switching", Jahrbücher für Nationalökonomie und Statistik 225(6), 653–674.

König, H. / *Wolters,* J. (1972): "Einführung in die Spektralanalyse ökonomischer Zeitreihen", Hain.

Kwiatkowski, D. / *Phillips,* P. C. / *Schmidt,* P. / *Shin,* Y. (1992): "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?", Journal of Econometrics 54(1), 159–178.

Lahiri, K. / *Wang,* J. G. (2013): "Evaluating Probability Forecasts for GDP Declines using Alternative Methodologies", International Journal of Forecasting 29(1), 175–190.

Leathwick, J. / *Elith,* J. / *Francis,* M. / *Hastie,* T. / *Taylor,* P. (2006): "Variation in Demersal Fish Species Richness in the Oceans Surrounding New Zealand: an Analysis using Boosted Regression Trees", Marine Ecology Progress Series 321, 267–281.

Lindlbauer, J. D. (1997): "Business Cycle Indicators", in: Business Cycle Indicators, ed. by Oppenländer, K.-H., chap. Beurteilung ausgewählter Frühindikatoren – Theorie und Evidenz, 144–176, Aldershot.

Liu, W. / *Moench,* E. (2014): "What Predicts US Recessions?", Staff Report No. 691, FRB of New York.

MacKinnon, J. G. (1996): "Numerical Distribution Functions for Unit Root and Cointegration Tests", Journal of Applied Econometrics 11(6), 601–618.

Malliaris, A. / *Malliaris,* M. (2015): "What Drives Gold Returns? A Decision Tree Analysis", Finance Research Letters 13, 45–53.

Manasse, P. / *Roubini,* N. (2009): ""Rules of Thumb" for Sovereign Debt Crises", Journal of International Economics 78(2), 192–205.

Marzban, C. (2004): "The ROC Curve and the Area under it as Performance Measures", Weather and Forecasting 19(6), 1106– 1114.

Mayr, A. / *Binder,* H. / *Gefeller,* O. / *Schmid,* M. (2014a): "The Evolution of Boosting Algorithms: From Machine Learning to Statistical Modelling", Methods of Information in Medicine 6(1), 419–427.

– (2014b): "Extending Statistical Boosting: An Overview of Recent Methodological Developments", Methods of Information in Medicine 6(2), 428–435.

McFadden, D. L. (1974): Conditional Logit Analysis of Qualitative Choice Behavior, Academic Press: New York, chap. 4. 105–142.

Mittnik, S. / *Robinzonov,* N. / *Spindler,* M. (2015): "Stock Market Volatility: Identifying Major Drivers and the Nature of Their Impact", Journal of Banking and Finance 58, 1 – 14.

National Bureau of Economic Research (2014): The NBER's Business Cycle Dating Committee, `http://www.nber.org/cycles/recessions.html`. (Last checked: 21.11.2014).

Ng, S. (2014): "Boosting Recessions", Canadian Journal of Economics 47(1), 1–34.

Nyberg, H. (2010)): "Dynamic Probit Models and Financial Variables in Recession Forecasting", Journal of Forecasting 29, 215 – 230.

OECD (2015): OECD Composite Leading Indicators: Reference Turning Points and Component Series, `http://www.oecd.org/std/leading-indicators/`. 06.10.2015.

Orphanides, A. / *Porter,* R. (2000): "P^* Revisited: Money-Based Inflation Forecasts with a Changing Equilibrium Velocity", Journal of Economics and Business 52(1/2), 87–100.

Pierdzioch, C. / *Rülke,* J.-C. (2015): "On the Directional Accuracy of Forecasts of Emerging Market Exchange Rates", International Review of Economics and Finance 38(4), 369–376.

Pierdzioch, C. / *Rülke,* J.-C. / *Tillmann,* P. (2015): "Using Forecasts to Uncover the Loss Function of FOMC Members", Macroeconomic Dynamics , forthcoming.

Proaño, C. R. / *Theobald,* T. (2014): "Predicting Recessions with a Composite Real-time Dynamic Probit Model", International Journal of Forecasting 30(4), 898–917.

R Core Team (2015): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, version 3.1.3, `http://www.R-project.org/`.

Raftery, A. R. (1995): "Bayesian Model Selection in Social Research", Sociological Methodology 25, 111–163.

Ridgeway, R. w. c. f. o. (2015): gbm: Generalized Boosted Regression Models. R package version 2.1.1. `http://CRAN.R-project.org/package=gbm`.

Robinzonov, N. / *Tutz,* G. / *Hothorn,* T. (2012): "Boosting Techniques for Nonlinear Time Series Models", AStA Advances in Statistical Analysis 96(1), 99–122.

Rudebusch, G. D. / *Williams,* J. C. (2009): "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve", Journal of Business and Economics Statistics 27(4), 492–503.

Savona, R. (2014): "Hedge Fund Systemic Risk Signals", European Journal of Operational Research 236, 282–291.

Savona, R. / *Vezzoli,* M. (2015): "Fitting and Forecasting Sovereign Defaults using Multiple Risk Signals", Oxford Bulletin of Economics and Statistics 77(1), 66–92.

Schapire, R. E. (2003): "The Boosting Approach to Machine Learning An Overview", in: Nonlinear Estimation and Classification, ed. by Denison, D. D. / Hansen, M. H. / Holmes, C. C. / Mallick, B. / Yu, B., *Lecture Notes in Statistics*, vol. 171, 149 – 171, Springer.

Schirwitz, B. (2009): "A Comprehensive German business Cycle Chronology", Empirical Economics 37(2), 287–301.

Schreiber, S. (2015): BreitungCandelonTest, version 1.5.1, `http://ricardo.ecn.wfu.edu/gretl/cgi-bin/current_fnfiles/BreitungCandelonTest.gfn`. Accessed: October 13, 2015.

Sephton, P. S. (1995): "Response Surface Estimates of the KPSS Stationarity Test", Economics Letters 47(3), 255–261.

Strobl, C. / *Malley,* J. / *Tutz,* G. (2009): "An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression trees, Bagging and Random Forests", Psychological Methods 14(4), 323–348.

Theobald, T. (2012): "Combining Recession Probability Forecasts from a Dynamic Probit Indicator", IMK at the Hans Boeckler Foundation, Macroeconomic Policy Institute.

Wohlrabe, K. / *Buchen,* T. (2014): "Assessing the Macroeconomic Forecasting Performance of Boosting: Evidence for the United States, the Euro Area and Germany", Journal of Forecasting 33(4), 231–242.

Wolters, J. (1996): "Zur Beurteilung verschiedener Frühindikatoren für die Produktionsentwicklung", in: Beiträge zur angewandten Wirtschaftsforschung: Festschrift für Karl Heinrich Oppenländer, ed. by Helmstädter, E. / Poser, G. / Ramser, H., 339–361, Duncker & Humblodt.

Wolters, J. / *Kuhbier,* P. / *Buscher,* H. S. (1990): "Die konjunkturelle Entwicklung in der Bundesrepublik Deutschland: Ergebnisse von Schätzungen alternativer aggregierter Konjunkturmodelle", Campus-Verlag.

Wolters, J. / *Lankes,* F. (1989): "Das Ifo-Geschäftsklima als Konjunkturindikator. Empirische Ergebnisse für die Verbrauchs- und Investitionsgüterindustrie", ifo Studien 35, 198–209.

Wuertz, D. (2013): fUnitRoots: Trends and Unit Roots. With contributions from many others (see the source file), version 3010.78, `http://CRAN.R-project.org/package=fUnitRoots`.

Youden, J., W (1950): "Index for Rating Diagnostic Tests", Cancer 3(3), 32–35.

Appendix

Table A1: Unit-Root Tests

Variable		ADF-GLS test		KPSS test	
	Start	Specification	Value	Specification	Value
		Levels			
Short-term interest rates	1960/1	c,12	-3.9325^{***}	c,6	2.8482^{***}
Spread	1960/1	c,12	5.4461^{***}	c,6	0.4418^{***}
Log of Order inflow	1960/1	c,t,12	1.3467	c,t,6	0.6630^{***}
Business climate	1960/1	c,12	-2.7524^{***}	c,6	0.4712^{***}
Business climate, expectations	1960/1	c,12	-2.1642^{**}	c,6	0.8411^{***}
Business climate, current situation	1961/1	c,12	-2.2678^{**}	c,6	0.6089^{***}
Log of consumer price index	1960/1	c,t,12	0.3898	c,t,6	2.2250^{***}
Log of share price index	1960/1	c,t,12	0.4103	c,t,6	0.7901^{***}
Log of industrial production	1960/1	c,t,12	1.4608	c,t,6	0.9722^{***}
Consumer confidence	1973/1	c,12	-2.3224^{**}	c,5	0.2147
Log of U.S. industrial production	1960/1	c,t,12	1.9838	c,t,6	0.8691^{***}
Log of Oil price	1960/1	c,t,12	0.3994	c,t,6	0.8488^{***}
OECD leading indicator	1961/1	c,12	-5.2996^{***}	c,6	0.0978
Log of Real effective exchange rate	1964/1	c,12	-2.7829^{*}	c,6	1.0441^{***}
U.S. short-term interest rates	1960/1	c,12	-1.8782^{***}	c,6	2.7350^{***}
		Change over previous period			
Short-term interest rates, mom	1960/2	c,12	-5.4279^{***}	c,6	0.0622
Spread, mom	1960/2	c,12	-5.6049^{***}	c,6	0.0131
Log of order inflow, yoy	1961/1	c,12	-4.7050^{***}	c,6	0.1048
Business climate, mom	1960/2	c,12	-7.2522^{***}	c,6	0.0271
Business climate. expectations, mom	1960/2	c,12	-6.1314^{***}	c,6	0.0199
Business climate. current situation	1961/2	c,12	-6.9391^{***}	c,6	0.0488
Log of consumer price index, yoy	1961/1	c,12	-1.9659^{**}	c,6	2.9269^{***}
Log of share price index, yoy	1961/1	c,12	-2.5115^{**}	c,6	0.1831
Log of industrial production, yoy	1961/1	c,12	-1.8099^{*}	c,6	0.2988
Consumer confidence, mom	1973/2	c,12	-1.5504	c,6	0.0567
Log of U.S. industrial production, yoy	1961/1	c,12	-1.1906	c,6	0.5506
Log of Oil price, yoy	1961/1	c,12	-4.3552^{***}	c,6	0.1099
OECD leading indicator, mom	1961/2	c,12	-6.0113^{***}	c,6	0.0113
Log of Real effective exchange rate, yoy	1965/1	c,12	-3.5045^{***}	c,6	0.0924
U.S. short-term interest rates, mom	1961/1	c,12	-6.2505^{***}	c,6	0.0832

Notes: *** (**, *) denotes rejection of the null hypotheses at the 1 (5, 10) % -level. ADF-GLS refers to the unit-root test proposed by Elliott et al. (1996). The critical values for the ADF-GLS-test are from MacKinnon (1996). KPSS refers to the test proposed by Kwiatkowski et al. (1992)."c" denotes that the model includes a constant, "t" denotes that the model includes a deterministic trend, and the number denotes how many lags of the endogenous variable are included in the model. mom denotes change over previous month, yoy denotes change over previous year. Computations have been undertaken with the R-package developed by Wuertz (2013).

Figure A1: Leading Indicators

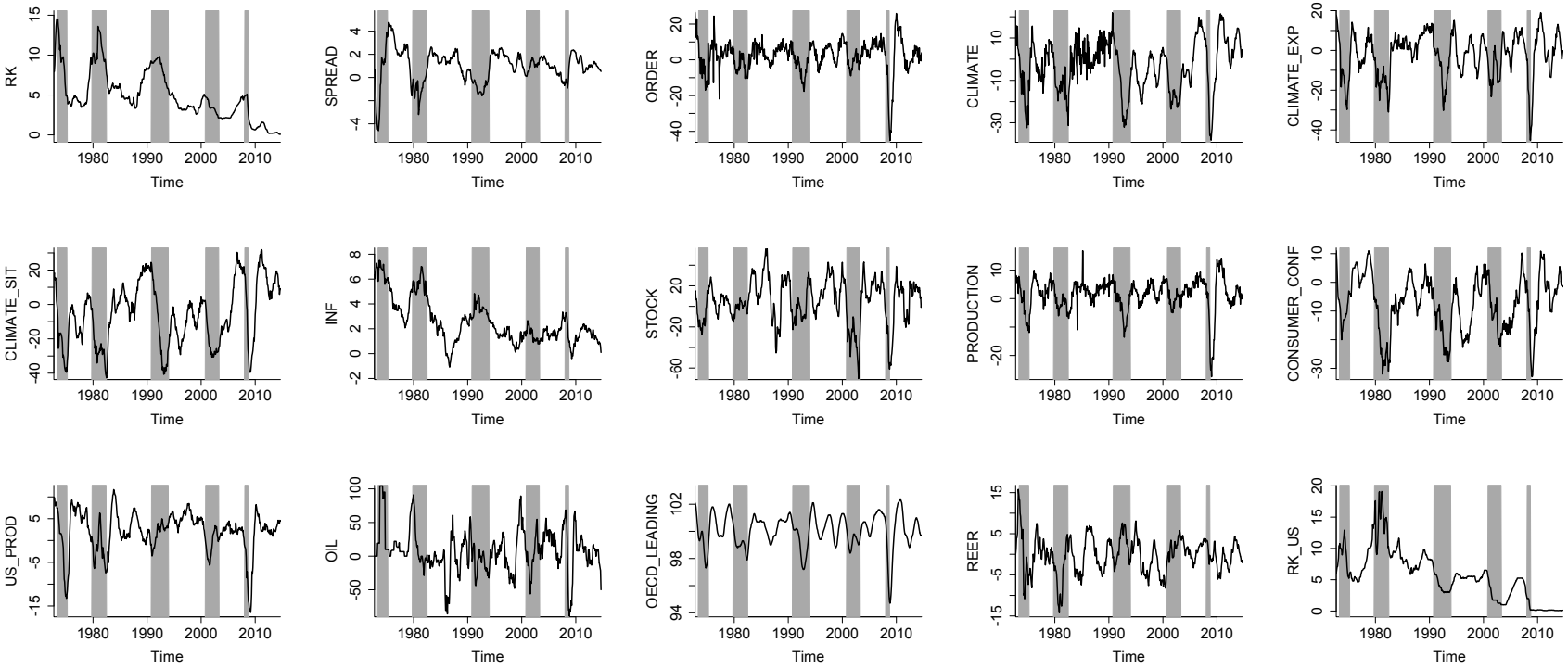
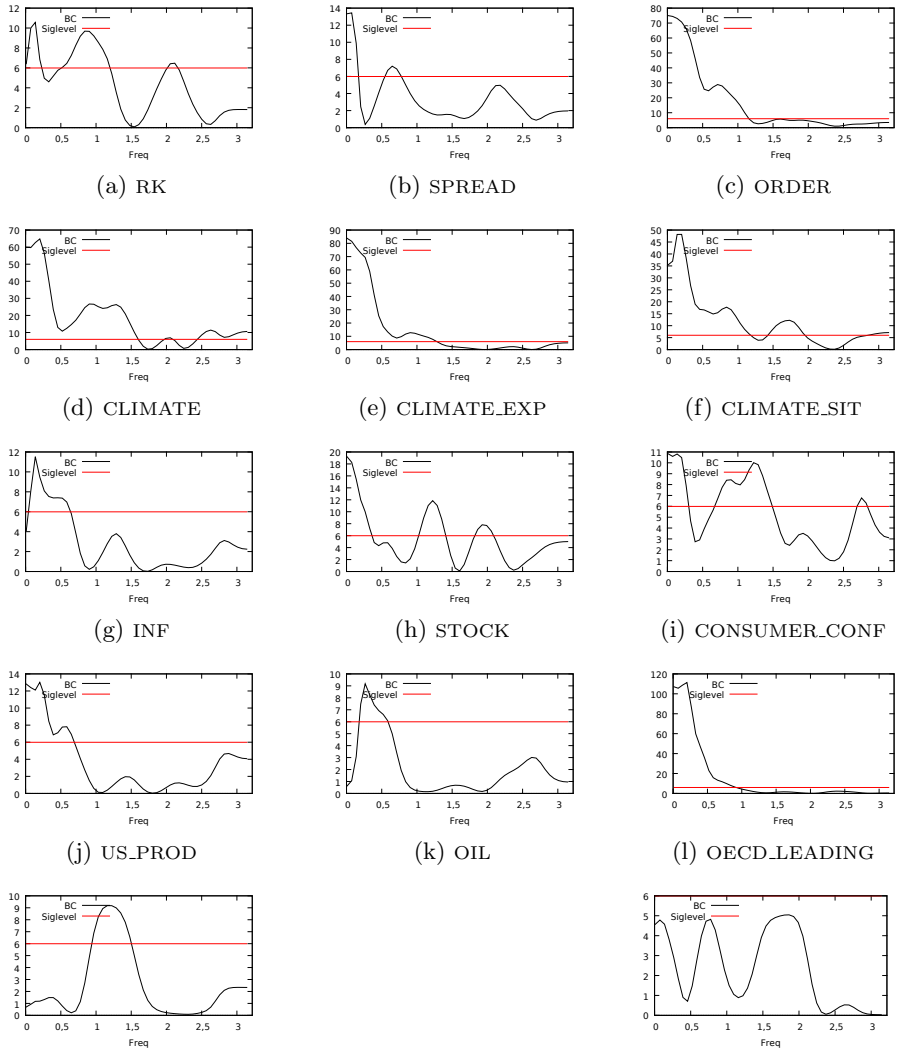


Figure A2: Results of the Breitung and Candelon (2006) Test



(a) RK (b) SPREAD (c) ORDER

(d) CLIMATE (e) CLIMATE_EXP (f) CLIMATE_SIT

(g) INF (h) STOCK (i) CONSUMER_CONF

(j) US_PROD (k) OIL (l) OECD_LEADING

(m) REER (n) RK_US

Note: The test was performed at each of 50 frequencies, equally spaced in the sequence $\{0, \pi\}$. The figures show the test values jointly with a 5% critical value (refers to the null hypothesis of no causality). The computations were performed using *gretl* (Cottrell and Lucchetti, 2015) and the package "BreitungCandelonTest" developed by Schreiber (2015).

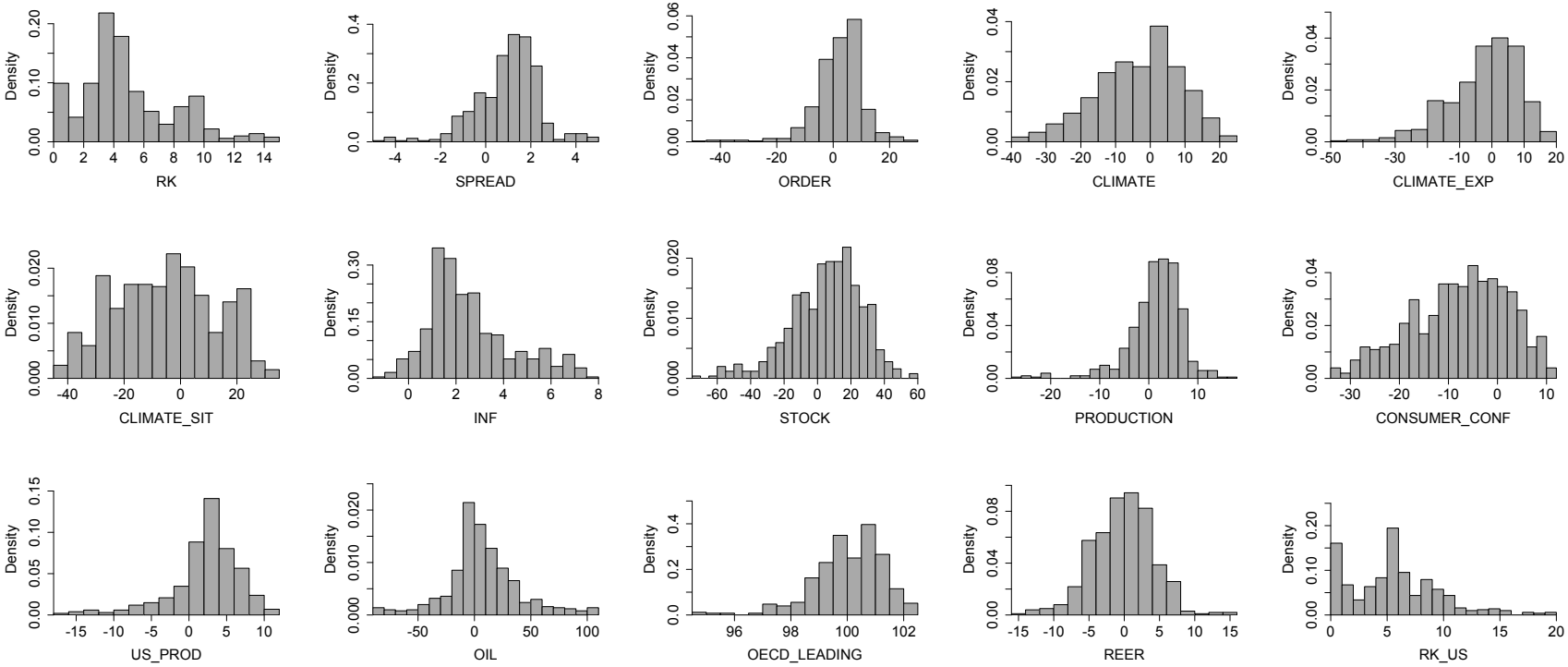



Figure A3: Histograms of the Leading Indicators