Predicting Recessions in Germany With Boosted Regression Trees

Jörg Döpke[1] Ulrich Fritsche[2] Christian Pierdzioch[3]

[1]University of Applied Sciences Merseburg

[2]Universität Hamburg and KOF ETH Zurich

[3]Helmut-Schmidt-Universität Hamburg

16th IWH-CIREQ Macroeconometric Workshop, December 2015

Motivation

- Reassessment of forecasting approaches and leading indicators after the "Great Recession"
- Poor performance of models and forecasters at the start of the "Great Recession" ▸ graph: Can't we do better than that?
- Things to do:
  1. Consider the bias-variance tradeoff!
  2. Consider a non-linear, state-dependent modelling approach!
  3. Consider appropriate evaluation metrics for binary outcomes!
  4. Consider the shape of the forecaster's/ general public's loss function before announcing a recession alarm!

Relationship to Existing Literature

- Probit models in recession predictions: Estrella and Hardouvelis (1991); Estrella and Mishkin (1998); for Germany see Fritsche and Kuzin (2005); Proaño and Theobald (2014)
- Regression trees: Used in empirical finance (Savona, 2014; Malliaris and Malliaris, 2015) and literature on financial crisis (Manasse and Roubini, 2009; Duttagupta and Cashin, 2011; Savona and Vezzoli, 2015)
- Boosting in economics: Berge (2015), Buchen and Wohlrabe (2011), Wohlrabe and Buchen (2014), Robinzonov et al. (2012), Bai and Ng (2009)
- Mostly related to our paper: Ng (2014)

Our Contribution

- In contrast to Ng (2014), we document marginal effects and interaction effects of the predictor variables. We also evaluate changing performance of the predictors over time.
- We employ ROC/AUROC statistics to evaluate the performance over all possible cut-off values.
- We relate the ROC/AUROC statistic to the loss function of the forecaster to define conditional optimal cut-offs for recession early warning systems.

Main results

- Short-term interest rates, the term spread and the stock market index are top leading indicators for recession probabilities. The importance changed over the time.
- The relationships are highly non-linear and interaction effects play a role.
- The BRT model significantly outperforms variants of the traditional probit model in pseudo out-of-sample comparisons.
- Attention: Under asymmetric loss, the optimal cut-off can be far below the 0.5 value.

Bias-Variance Tradeoff

The Methodological Challenge

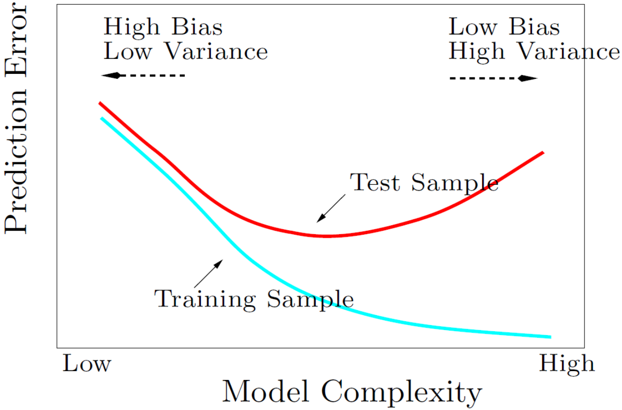


Figure : Bias-Variance-Tradeoff, figure adopted from Hastie et al. (2009), p. 38.

Boosting as a Step-by-Step Modeling Device

- Loss function:

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

with $\tilde{y}_{t+k} = 2y_{t+k} - 1$, such that $\tilde{y}_{t+k} \in \{-1, 1\}$.

- First-order condition:

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

- Construct $F(\mathbf{x_t})$ as a sum of simpler functions

$$F(\mathbf{x_t}) = \sum_{m=0}^{M} T_m(\mathbf{x_t}), \tag{3}$$

Boosting as a Step-by-Step Modeling Device

Gradient-descent boosting (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:
   - 3.1 Compute the negative gradient vector given by $z_{t,m} = -\partial \mathcal{L}(F)/\partial F$
   - 3.2 Fit a weak learner, $T_m(\mathbf{x_t})$, to the negative gradient vector.
   - 3.3 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 (3).
   - 3.4 Equipped with the new function estimate, go back to Step (3.1).
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 $m = 0, ..., M$.

Regression Trees as Weak Learners

- A regression tree (for nontechnical introduction, see Hastie et al. (2009)), $T(\mathbf{x_t})$, 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.

Regression Trees as Weak Learners

Basic Principle

- Linear regression model vs. regression tree

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad\qquad y \sim T(x_1, x_2)$$

- Define all partitions

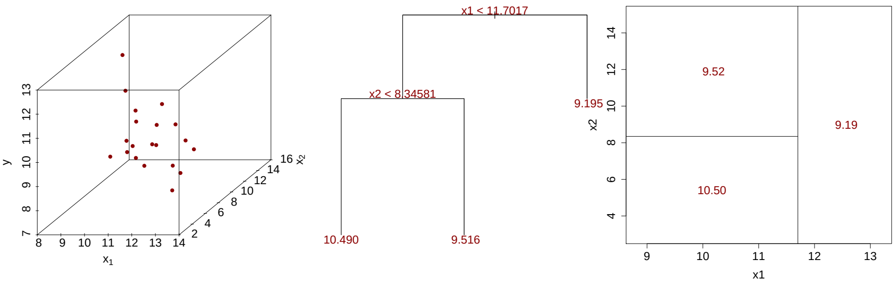
$$R_1(s,p) = \{x_s | x_s \leq p\} \text{ and } R_2(s,p) = \{x_s | x_s > p\}$$
$$s = 1, 2 \text{ and } p = \text{partitioning point}$$

- Find the optimal partition

$$\min_{s,p} \left\{ \min_{\bar{y}_1} \sum_{x_s \in R_1(s,p)} (y_i - \bar{y}_1)^2 + \min_{\bar{y}_2} \sum_{x_s \in R_2(s,p)} (y_i - \bar{y}_2)^2 \right\}$$

Regression Trees as Weak Learners

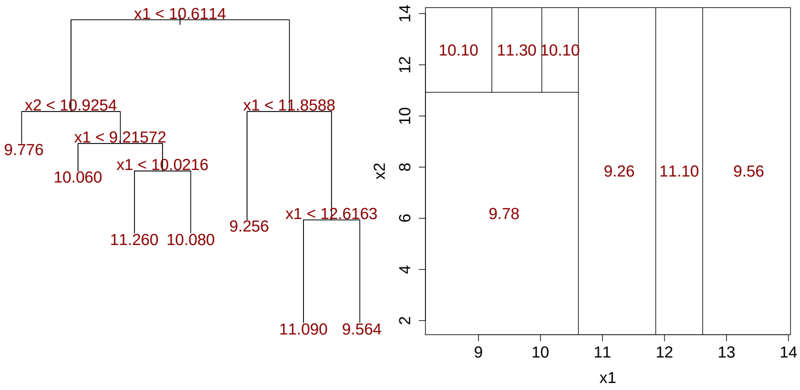
Example, $N = 20$



$$\hat{y}_i = T\left(\mathbf{x}_i, \{R_l\}_1^3\right) = \sum_{l=1}^{3} \bar{y}_l \mathbf{1}(\mathbf{x}_i \in R_l), \quad i = 1, ..., 20, \quad \mathbf{x}_i = (x_{1,i}, x_{2,i})$$

Regression Trees as Weak Learners

Example, $N = 50$



Regression Trees as Weak Learners

- In step 3.1 of the gradient boosting: choose optimal terminal node responses for the given loss function by applying Newton's method such that:

$$\gamma_{l,m} = \arg\min_{\gamma} \sum_{\mathbf{x_t} \in R_{l,m}} \mathcal{L}(F_{m-1}(\mathbf{x_t}) + \gamma). \tag{4}$$

- Function $F(\mathbf{x_t})$ takes the form:

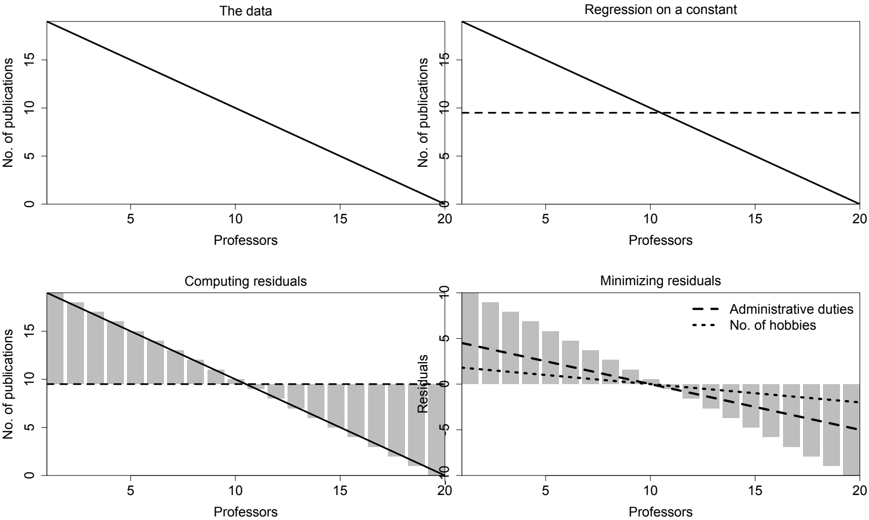
$$F(\mathbf{x_t}) = \sum_{m=0}^{M} \gamma_{l,m} \mathbf{1}_{\mathbf{x_t} \in R_{l,m}} \tag{5}$$

Regression Trees as Weak Learners

- 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.

$$F_m(\mathbf{x_t}) = F_{m-1}(\mathbf{x_t}) + \lambda \gamma_{l,m} \mathbf{1}_{\mathbf{x_t} \in R_{l,m}}. \tag{6}$$

Boosting – A Simple Example



Evaluating Forecasts

Receiver Operating Characteristics

- c = cutoff:

$$\hat{y}_{t+k}(c) = \begin{cases} 1, & \text{if} \quad \hat{P}(y_{t+k} = 1|\mathbf{x_t}) \geq c, \\ 0, & \text{if} \quad \hat{P}(y_{t+k} = 0|\mathbf{x_t}) < c. \end{cases} \tag{7}$$

$$PTP(c) = \frac{1}{n_R} \sum_{t=1}^{N} \mathbf{1}_{y_{t+k} = \hat{y}_{t+k} = 1} \tag{8}$$

$$PTN(c) = \frac{1}{n_{NR}} \sum_{t=1}^{N} \mathbf{1}_{y_{t+k} = \hat{y}_{t+k} = 0} \tag{9}$$

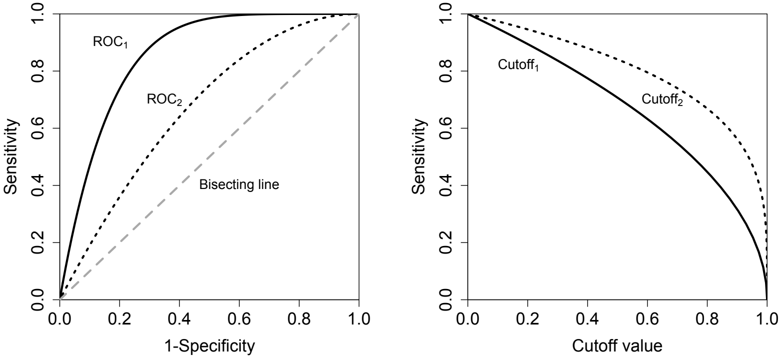
- $AUROC$ = area under the ROC curve
- Non-parametric approach:[1]

$$AUROC = \frac{n_{NR} n_R - U}{n_{NR} n_R}, \tag{10}$$

[1] U = Wilcoxon-Mann-Whitney U statistic (Bamber, 1975).

Evaluating Forecasts

Receiver Operating Characteristics



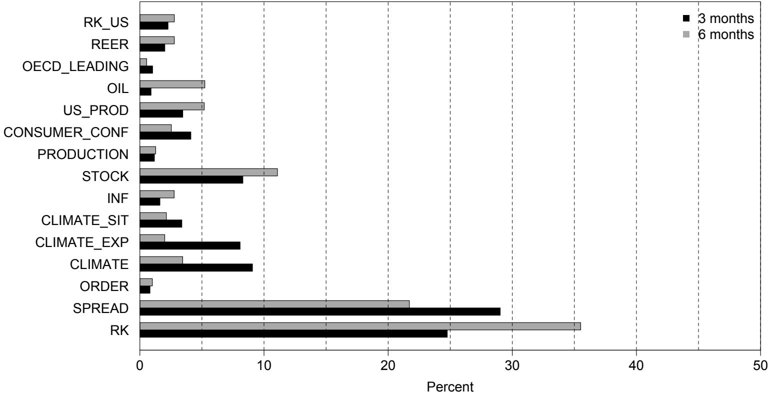
Data

Table : Data & Sources

Series	Source	ID
Recession phases	Economic Cycle Research Institute	RECESSION
Short term interest rates	Deutsche Bundesbank	RK
Interest rate spread	Deutsche Bundesbank	SPREAD
Order inflow industry	Deutsche Bundesbank	ORDER
ifo business climate	FRED / ifo institute	CLIMATE
ifo business climate, current situation	FRED / ifo institute	CLIMATE_SIT
ifo business climate, expectations	FRED / ifo institute	CLIMATE_EXP
Consumer price index (CPI)	Deutsche Bundesbank	INF
OECD Stock Market Index	OECD Monthly Economic Indicators	STOCK
Industrial production	Deutsche Bundesbank	PRODUCTION
OECD Consumer Confidence	OECD Monthly Economic Indicators	CONSUMER_CONF
U.S. Industrial production	OECD Monthly Economic Indicators	US_PROD
Crude Oil Prices: West Texas Intermediate (WTI)	FRED	OIL
OECD leading indicator	OECD Monthly Economic Indicators	OECD_LEADING
Real Narrow Effective Exchange Rate for Germany	FRED	REER
U.S. Effective Federal Funds Rate	FRED	RK_US

graph

Relative Importance of Leading Indicators



Marginal Effects

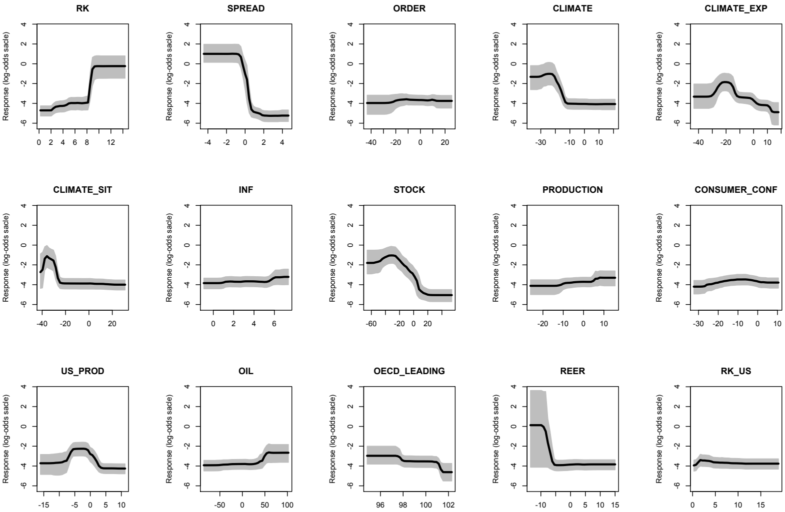


Figure : Marginal effects, horizon = 3

Marginal Effects

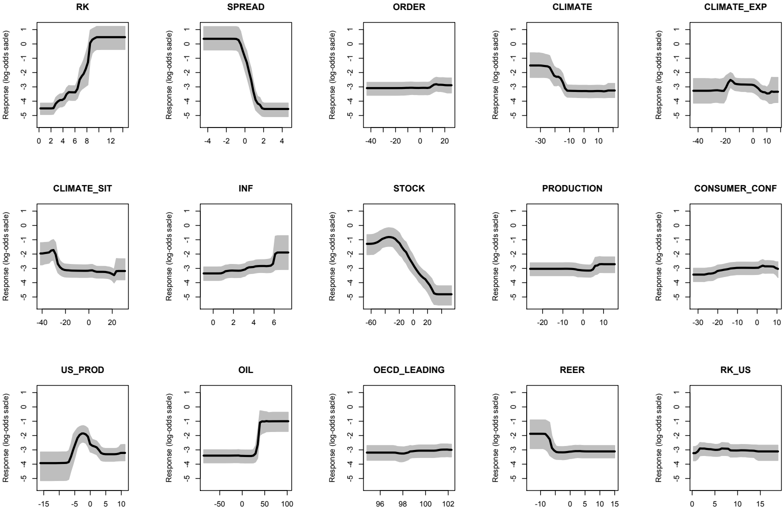
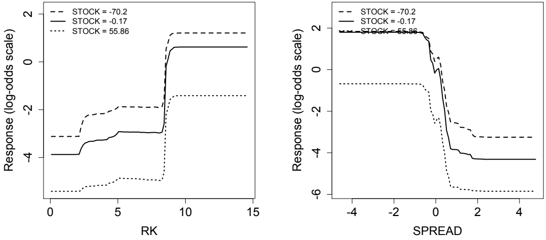


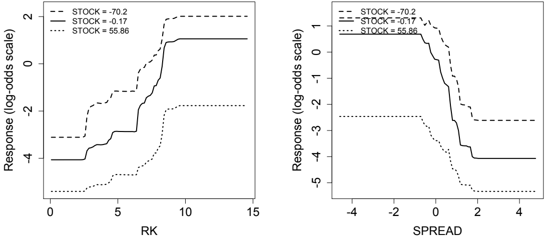
Figure : Marginal effects, horizon = 6

Interaction Effects: Influence of the Stock Market

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months



ROC analysis

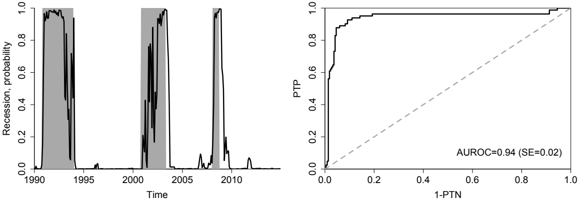
Table : AUROC Statistics for the BRT Approach

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

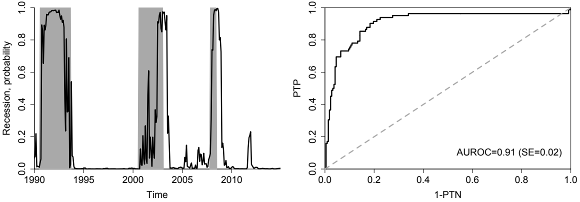
Note: CI = 95% confidence interval. Test fraction: 30%. Number of simulation runs: 1,000. For the comparisons of the BRT approach with the Probit approach, mean denotes the average difference between the AUROC statistic for the BRT approach minus the AUROC statistic for the Probit approach.

Recursive Estimation: Out-of-Sample

Panel A: Forecast Horizon: 3 Months

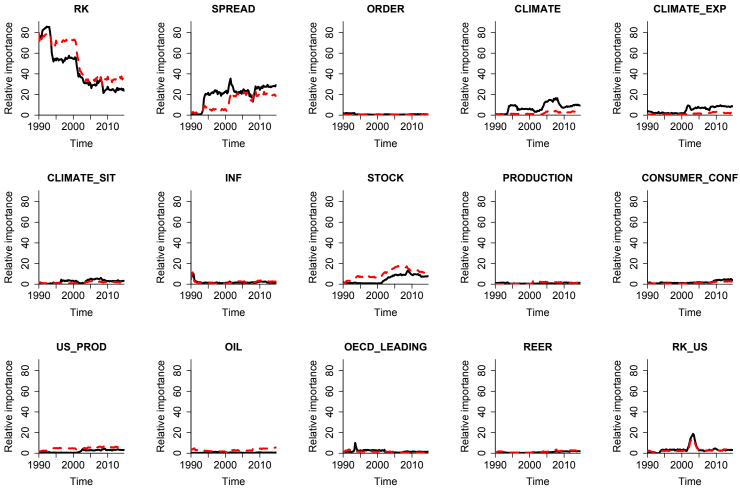


Panel B: Forecast Horizon: 6 Months



Variation Across Time

Figure : Changing Relative Importance of Leading Indicators

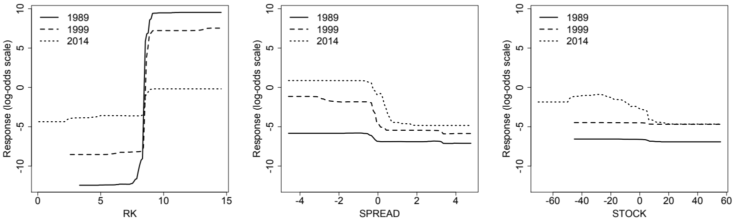


Note: Black (red) solid (dashed) lines: Forecast horizon 3 (6) months.

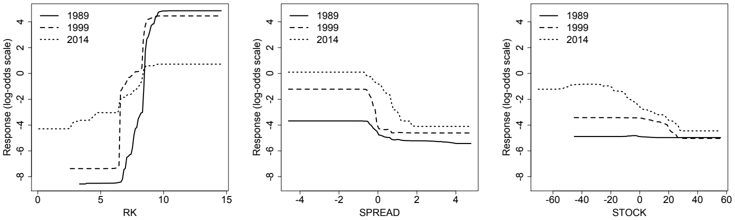
Variation Across Time

Figure : Changing Marginal-Effect Curves

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months



Optimal Cutoff Value

- Optimal cutoff value: 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)
- Symmetric loss function: maximizing utility is equivalent to maximizing the efficiency, E of forecasts defined as $E(c) = \pi PTP(c) + (1-\pi)PTN(c)$

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

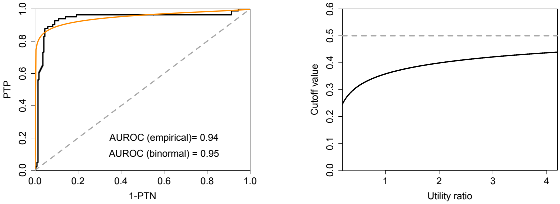
Index / Forecast horizon	3 months	6 months
Efficiency	0.9682	0.9514
Cutoff value	0.3604	0.3303

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

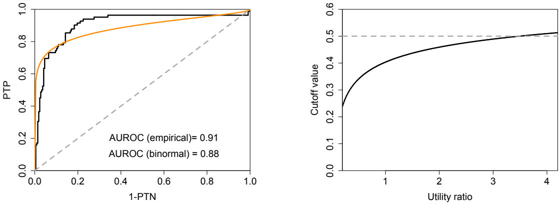
Optimal Cutoff Value

Figure : Asymmetric Loss and the Optimal Cutoff Value

Panel A: Forecast Horizon: 3 Months



Panel B: Forecast Horizon: 6 Months



Conclusion

- BRT as one useful tool to construct highly efficient "early warning" systems
- Method beats simple Probit models as well as BMA Probit models easily
- Non-linearities and interaction effects are modeled simultanously and this might give further insights into the process leading the economy into recessions
- To-Do's
  - Dynamic probit models and real time data?
  - Big data?

Thank you for your attention!

The Misfortune of Forecasters and Models

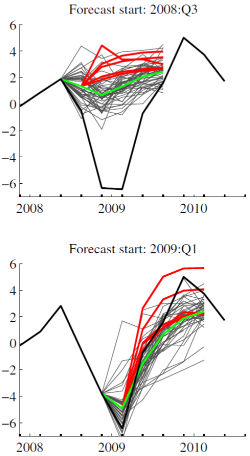
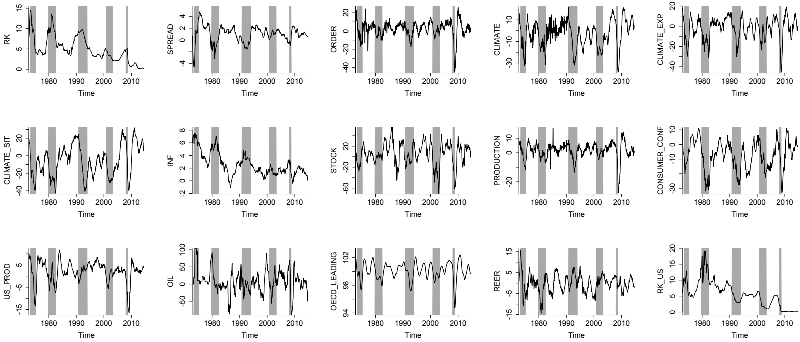


Figure : Performance of SPF Forecasters and Models, adopted from Wieland and Wolters (2013), p. 303

Note: Gray and Green lines show individual and consensus forecasts, red lines indicate model-based forecasts back

Data


RK
SPREAD
ORDER
CLIMATE
CLIMATE_EXP
CLIMATE_SIT
INF
STOCK
PRODUCTION
CONSUMER_CONF
US_PROD
OIL
OECD_LEADING
REER
RK_US
Time
1980
1990
2000
2010

back

Robustness

- Drop three variables and use the sample 1961:M1 to 2014:12, apply 50% training sample / 50% test sample split
- Apply different evaluation metrics
- General results remain the same!

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

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.

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

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.

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.

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.

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

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

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.

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

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

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

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.

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.