Financial Crisis Prediction with Machine Learning

Modelling with Big Data and Machine Learning
26 November, 2018

Kristina Bluwstein
Marcus Buckmann
Andreas Joseph
Miao Kang
Bank of England

Sujit Kapadia
European Central Bank

Özgür Şimşek
University of Bath

Research goals

Develop a prediction model for financial crises in developed countries that

... predicts crises ahead of time

... is accurate

... transparent and easy to communicate

... provides a continuous risk measure

Research goals

Develop a prediction model for financial crises in developed countries that

... predicts crises ahead of time

... is accurate

... transparent and easy to communicate

... provides a continuous risk measure

To achieve that we

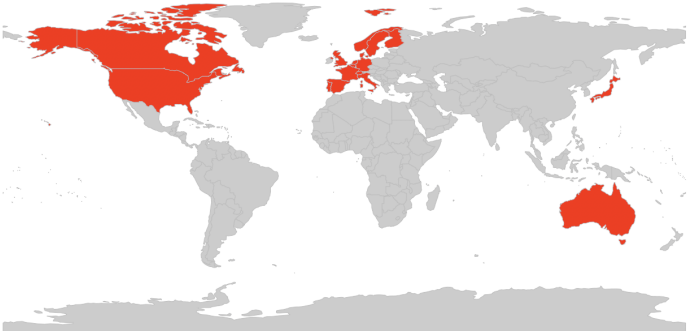
... conducted a horse race of machine learning models

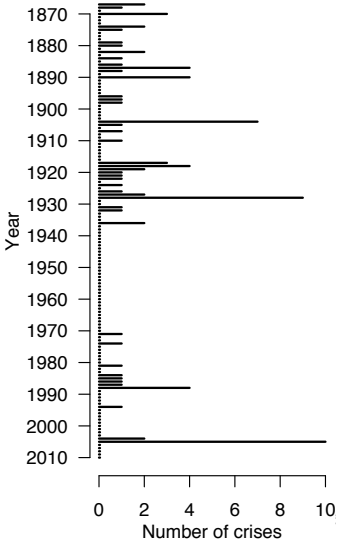
... made their predictions transparent using Shapley values

Jordà-Schularick-Taylor Macrohistory Database

Observations

- 17 developed countries, annual data between 1870 and 2013
- 92 crisis, 2356 non-crisis observations
- 24 potential indicators

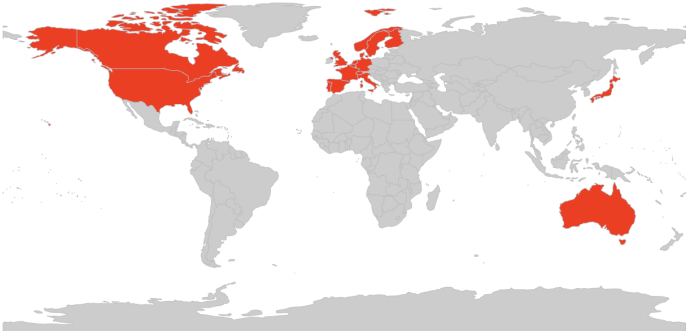




Jordà-Schularick-Taylor Macrohistory Database

Observations

- 17 developed countries, annual data between 1870 and 2013
- 92 crisis, 2356 non-crisis observations
- 24 potential indicators



Variables we use

- Total credit
- Slope of the yield curve
- Debt service ratio
- Current account
- Stock Prices
- CPI
- Consumption
- Investment
- Broad money
- Public debt

Problem dimensions

Transformation selection (e.g. Percentage changes vs. ratio changes)

↓

Variable selection (e.g. Long- short term rate vs. long and short term rate)

↓

Horizon of change (growth rates/ratio changes of 1,..,5 years)

↓

Landing zone (Predict 1 or 1–2 years before a crisis)

↓

Model selection (Horse race of machine learning models)

Baseline dataset

- Make data stationary
  - ◇ 2-year ratio change: $\frac{credit_{1980}}{GDP_{1980}} - \frac{credit_{1978}}{GDP_{1978}}$
  - ◇ 2-year growth rates: $\frac{stocks_{1980} - stocks_{1978}}{stocks_{1978}}$
- Predict one and two years in advance of a crisis.
- Exclude actual crisis observation and five following years (post-crisis bias).
- Exclude world wars and 1933–1938
- Compute global credit growth: Mean credit to GDP growth across all countries

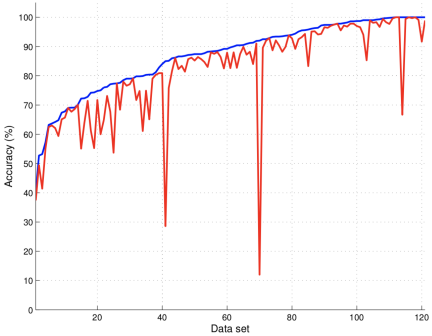
Which model should we use?

- Fernández-Delgado et al. (2014) tested 179 classification models on 121 data sets.
- Random forest implementations on ranks 1,2, and 5.
- Support vector machines (SVM) on ranks 3 and 4.

Which model should we use?

- Fernández-Delgado et al. (2014) tested 179 classification models on 121 data sets.
- Random forest implementations on ranks 1,2, and 5.
- Support vector machines (SVM) on ranks 3 and 4.

There is no free lunch!



Best model
Random forest

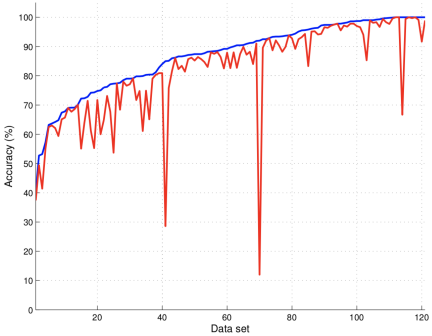
Source: Fernández-Delgado et al. (2014)

Which model should we use?

- Fernández-Delgado et al. (2014) tested 179 classification models on 121 data sets.
- Random forest implementations on ranks 1,2, and 5.
- Support vector machines (SVM) on ranks 3 and 4.

→ **Horse race needed!**

There is no free lunch!

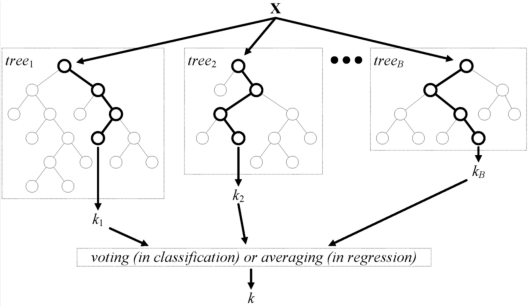


Best model
Random forest

Source: Fernández-Delgado et al. (2014)

Tree-based models

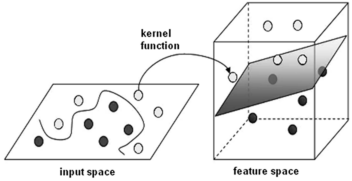
- Decision tree
  - ◇ Easy to interpret
  - ◇ Risk of overfitting
- Random forest
  - ◇ Collection of diverse decision trees
  - ◇ Less overfitting
- Extreme trees
  - ◇ Extension of random forest



Source: Verikas et al. (2016)

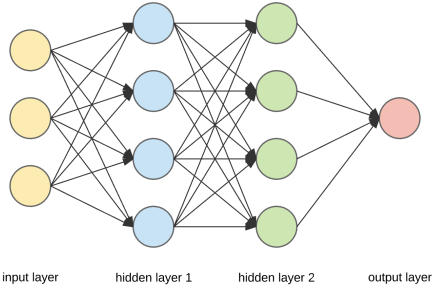
Other models

Support vector machines



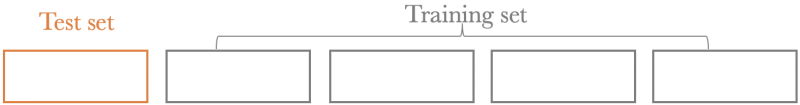
Source: Zararsiz and Coşgun (2014)

Neural networks

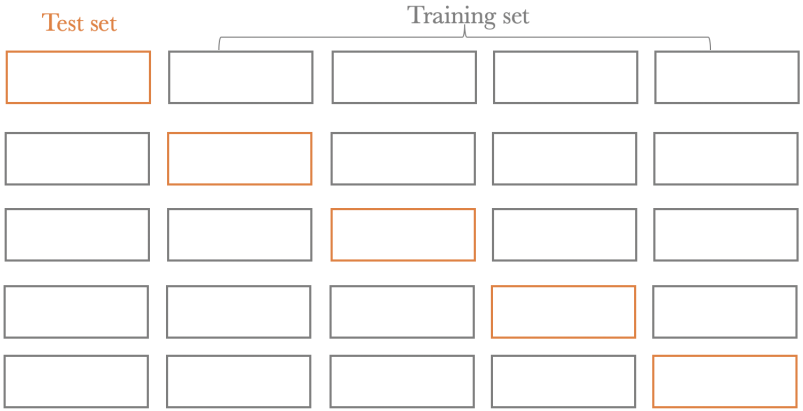


Source: towardsdatascience.com

Cross-validation to estimate out-of-sample performance



Cross-validation



100 x 5-fold cross-validation

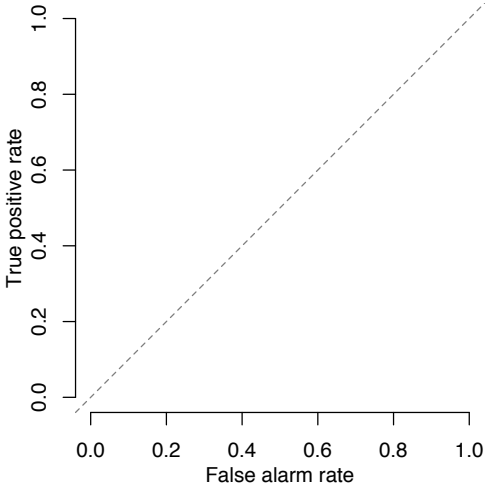
Nested cross-validation to learn hyperparameters.

Grid search:

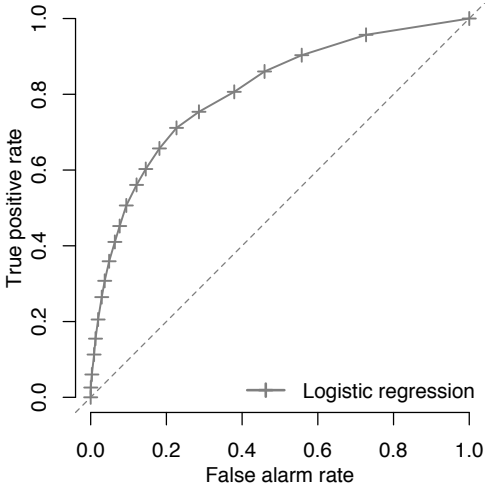
C	10^{-5}	10^{-5}	10^{-5}	...	10^{4}
γ	10^{-5}	10^{-4}	10^{-3}	...	10^{4}

$\Rightarrow$ Computationally very costly!

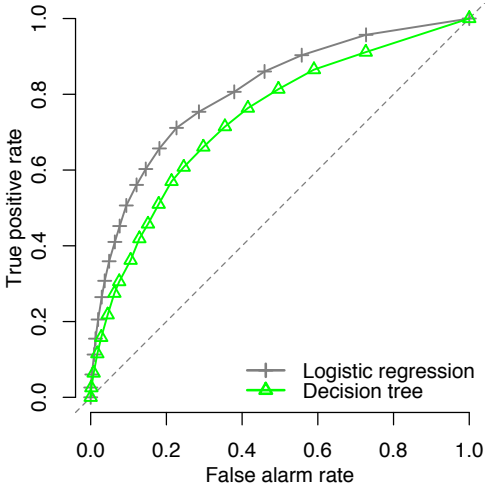
Out-of-sample performance in the ROC space



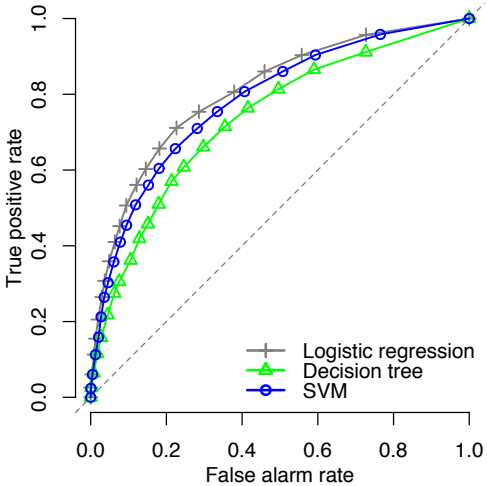
Linear baseline



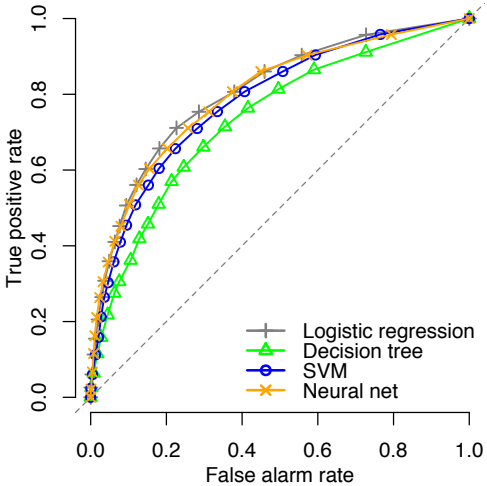
+ Decision trees



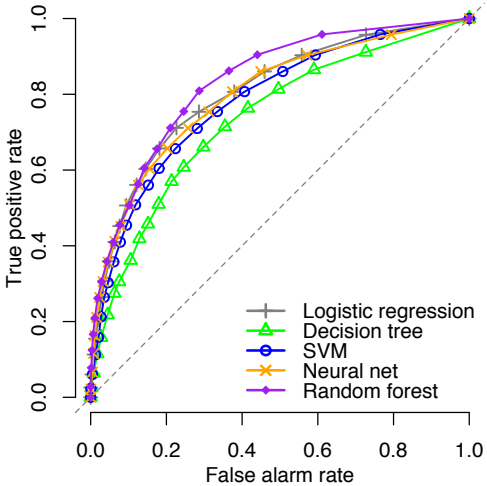
+ SVM


True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False alarm rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic regression
Decision tree
SVM

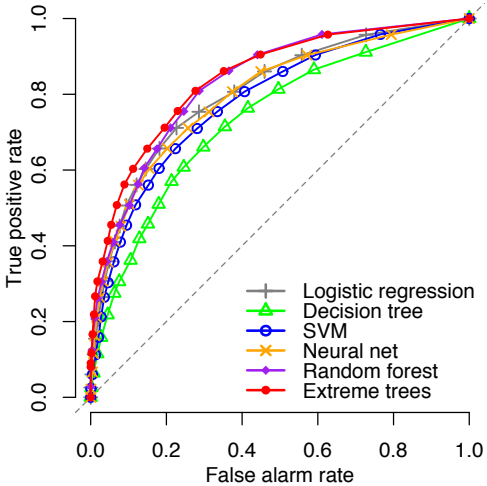
+ Neural network



+ Random forest



The winner is Extreme trees



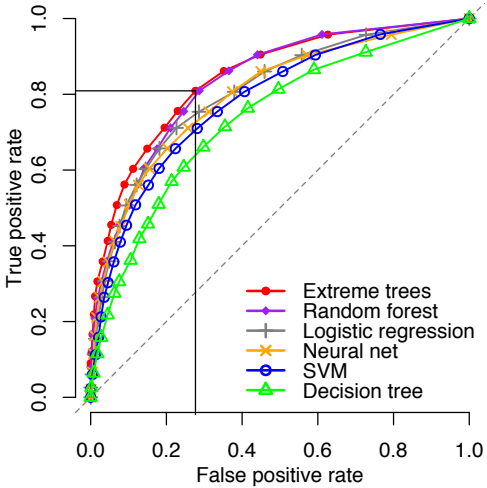
Area under the curve (AUC) performance

Extreme trees	0.84
Random forest	0.83
Logistic regression	0.80
Neural net	0.80
SVM	0.79
Decision tree	0.73

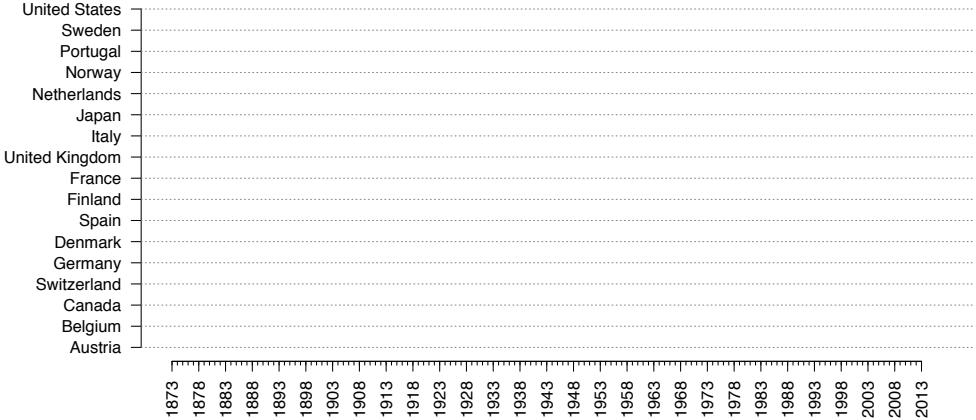
100 replications of 5-fold cross-validation.

Standard errors not shown but consistently below 0.002.

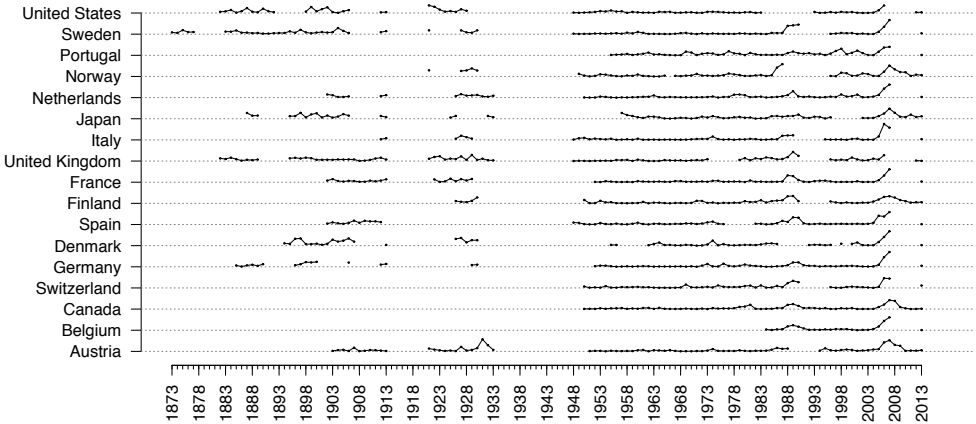
How well can we predict individual observations?



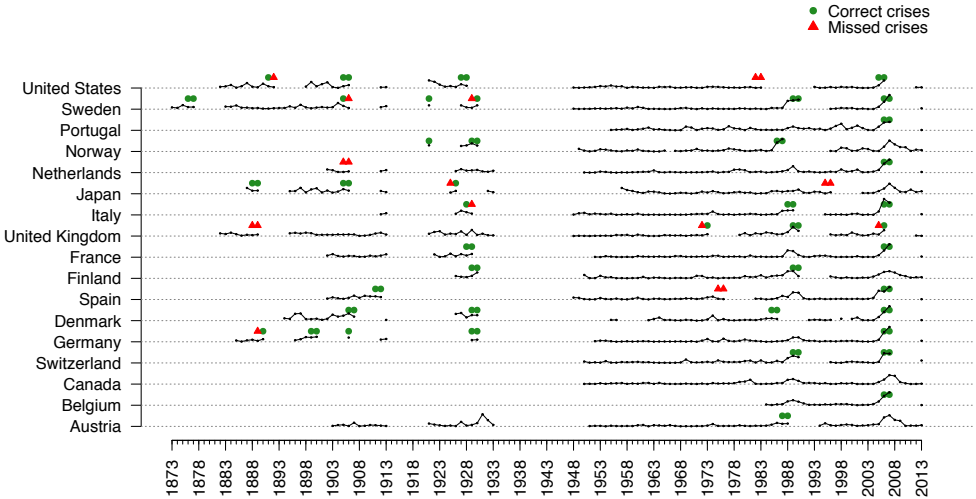
Predicting individual observations (Extreme trees)



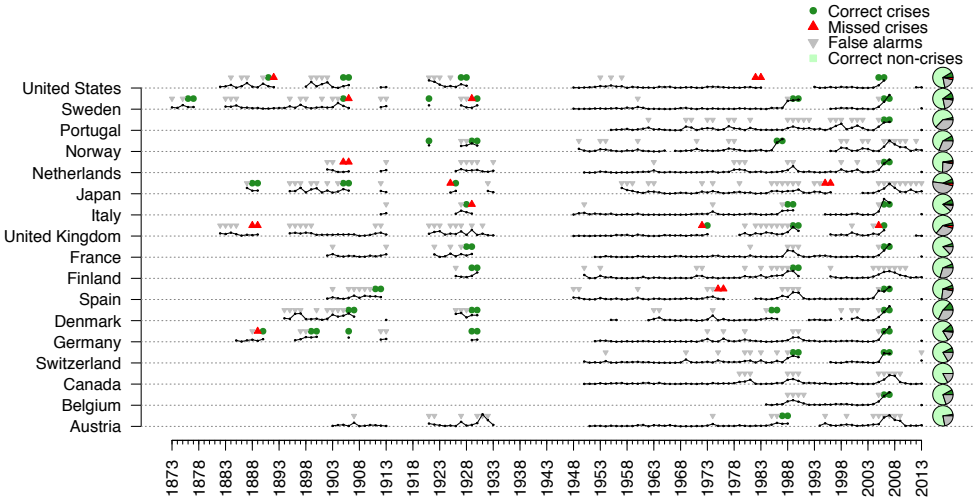
Predicted values



We only miss 5 of 48 crises completely



Many false alarms – but many more correct non-crises predictions



Shapley values in Game Theory

- How much does player A contribute a collective payoff f obtained by a group N? (Shapley, 1953).
- Observe payoff of the group with and without player A.
- Contribution depends on the other players in the game.
- All possible coalitions need to be evaluated.

$$\phi_j = \sum_{S \subseteq N \setminus j} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\} - f(S)]$$

$2^{|N|-1}$ coalitions are evaluated.
Computationally complex!

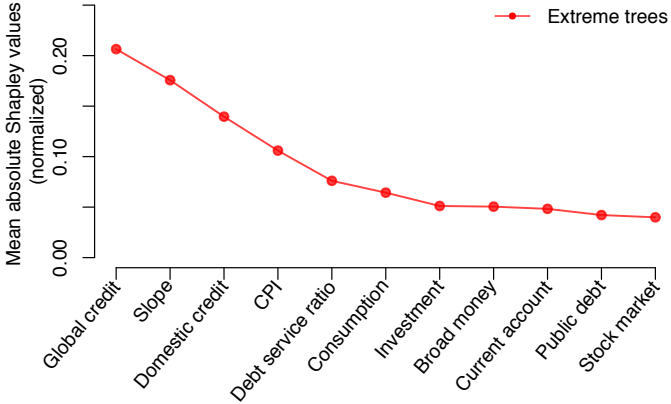
Shapley values applied to our problem

	Game Theory	Machine Learning
N	Players	Predictors
f	Collective payoff	Predicted value for one observation
S	Coalition	Predictors used for prediction
Source	Shapley (1953)	Strumbelj and Kononenko (2010)
		Lundberg and Lee (2017)

Shapley values of observation i sum up to predicted value: $\hat{y}_i = \sum_{j=1}^{k} \phi_{ij} + c$.

Toy example

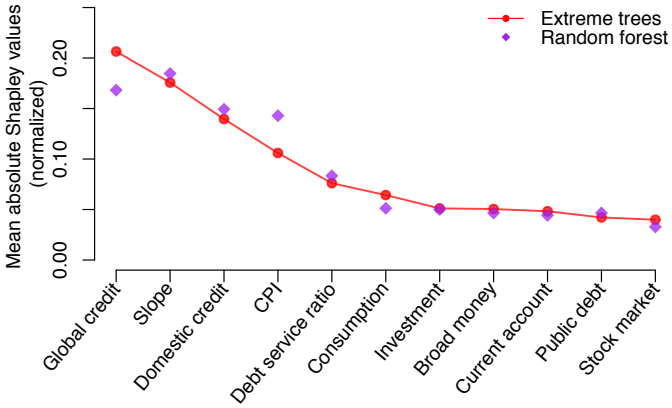
Importance of predictors across all observations



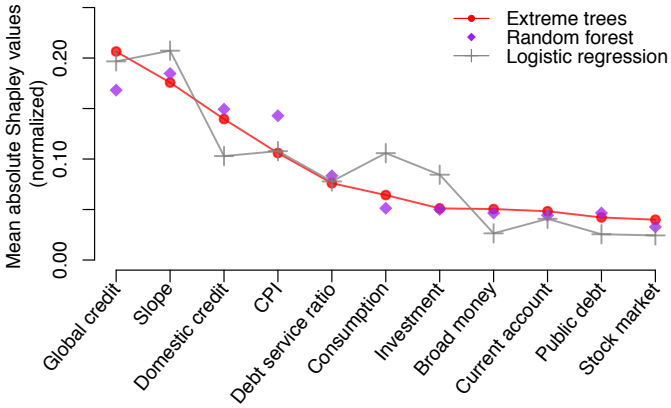
Contributions to the literature:

- Our results confirm: Global and domestic credit are important
- New results: The slope of the yield curve is important
- Current account and money are less important

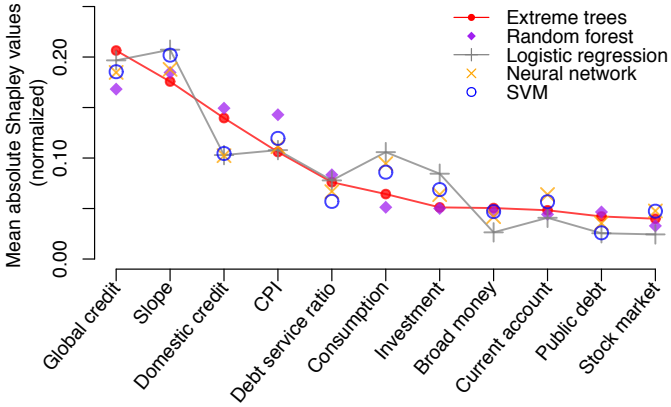
High agreement across the models



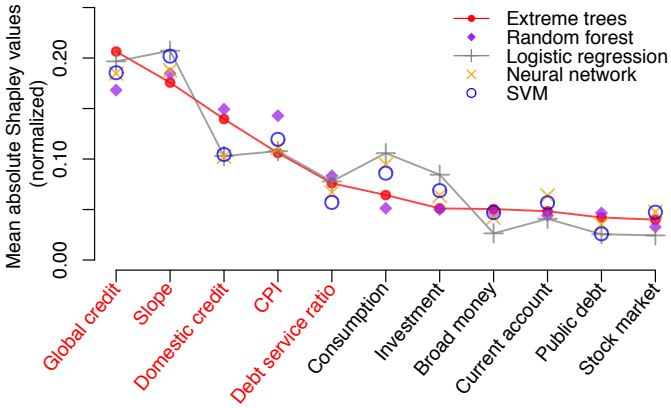
High agreement across the models



High agreement across the models



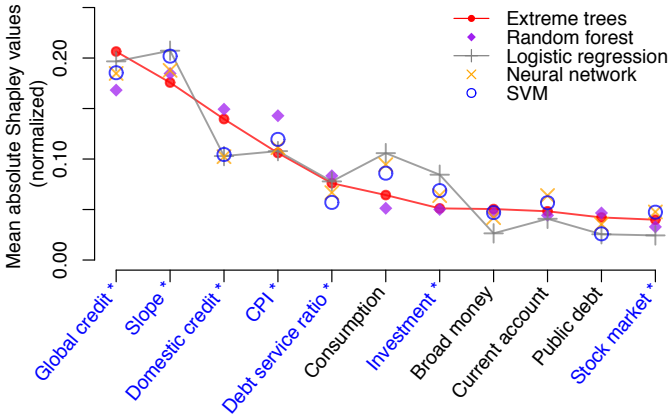
Top five predictors suffice for prediction



Mean AUC:

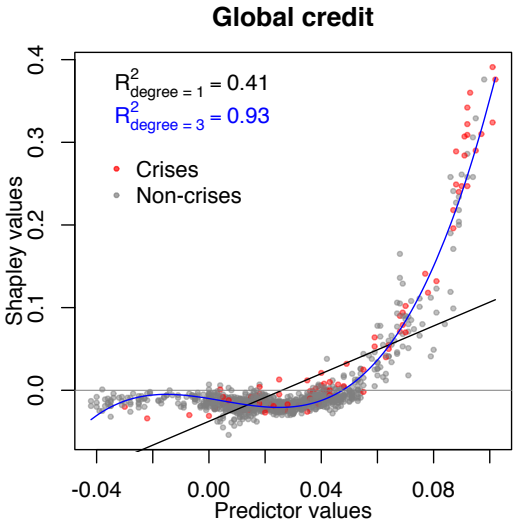
	Top five	All
Extreme trees	0.84	0.84
Random forest	0.84	0.83
Log. regression	0.80	0.80
Neural net	0.81	0.80
SVM	0.81	0.79
Decision tree	0.75	0.73

Shapley regression (Joseph, 2018)



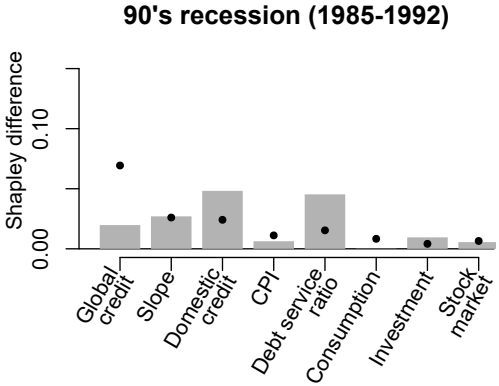
- Statistical test of Shapley values
- Regressing crisis indicator on Shapley values

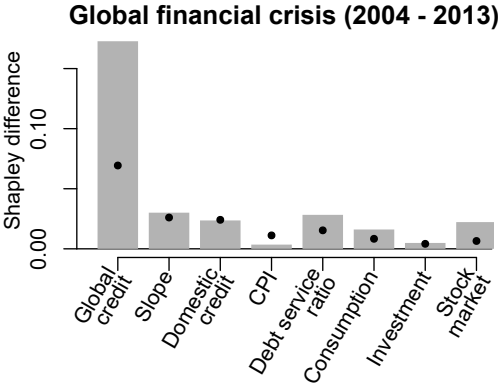
Nonlinearity of extreme trees



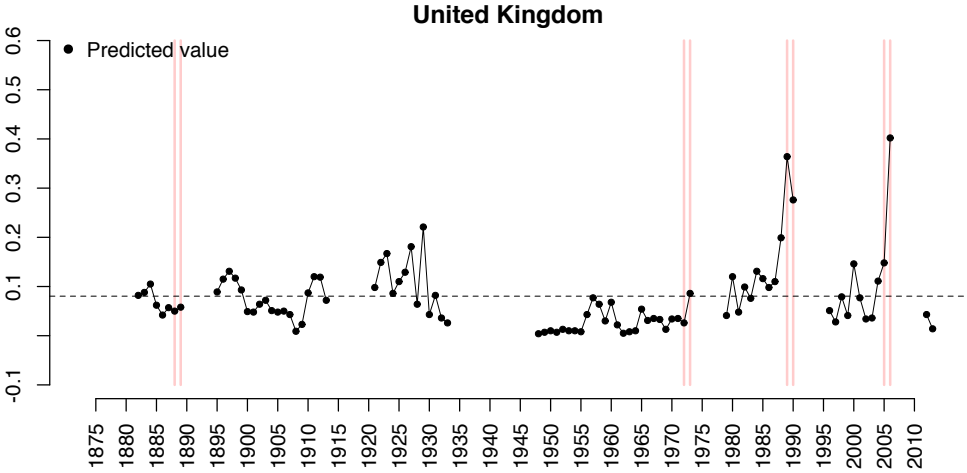
- No unexpected nonlinearities
- Directions of associations match the linear model

Change of Shapley values over time

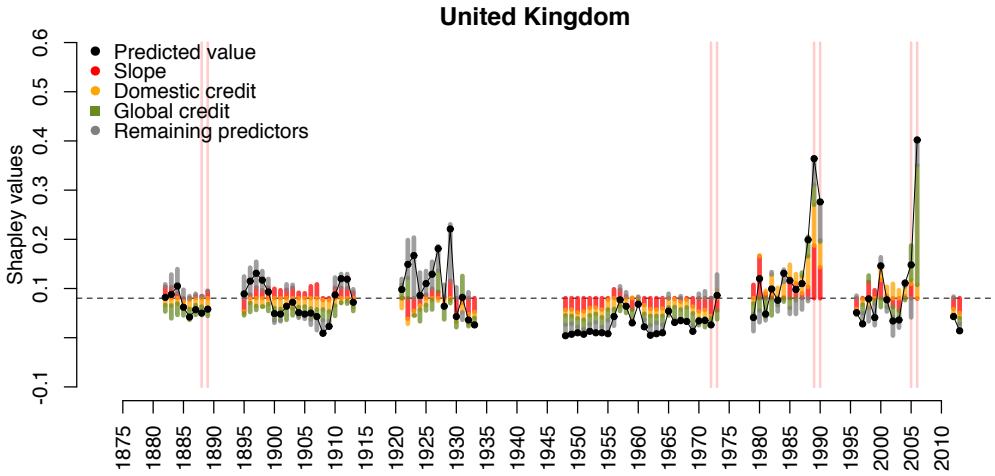




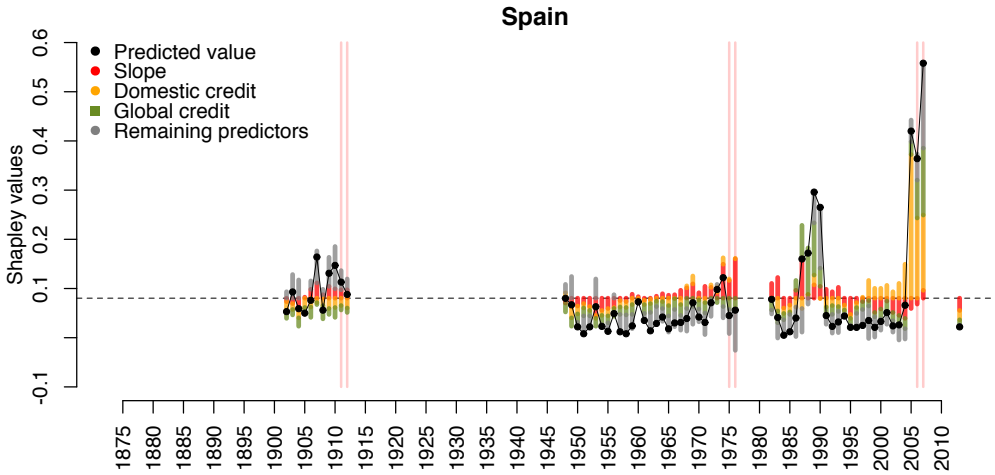
Model decomposition for individual observations



Model decomposition for individual observations



Model decomposition for individual observations



Summary

- Tree based machine learning models outperform logistic regression
- Predictions of black box models are surprisingly simple functions of the predictors
- Credit (domestic and global) and the slope of the yield are key indicators

Work in progress / Future work

- Understand interactions of variables
- Develop a simple decision aid predicting crises (e.g. a small tree)
- Develop economic intuition behind main predictors

References I

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 15(1):3133–3181.

Fujimoto, K., Kojadinovic, I., and Marichal, J.-L. (2006). Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55(1):72–99.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.

Joseph, A. (2018). Shapley regression: A tool to position machine learning within the spectrum of econometric and policy models. Submitted to as a Bank of England Staff Working Paper.

References II

Liu, W. and Moench, E. (2016). What predicts US recessions? *International Journal of Forecasting*, 32(4):1138–1150.

Lundberg, S. M., Erion, G. G., and Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, pages 4765–4774.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Strumbelj, E. and Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18.

References III

Verikas, A., Vaiciukynas, E., Gelzinis, A., Parker, J., and Olsson, M. C. (2016). Electromyographic patterns during golf swing: Activation sequence profiling and prediction of shot effectiveness. *Sensors*, 16(4):592.

Zararsiz, G. and Coşgun, E. (2014). Introduction to statistical methods for microrna analysis. In *miRNomics: MicroRNA Biology and Computational Analysis*, pages 129–155. Springer.