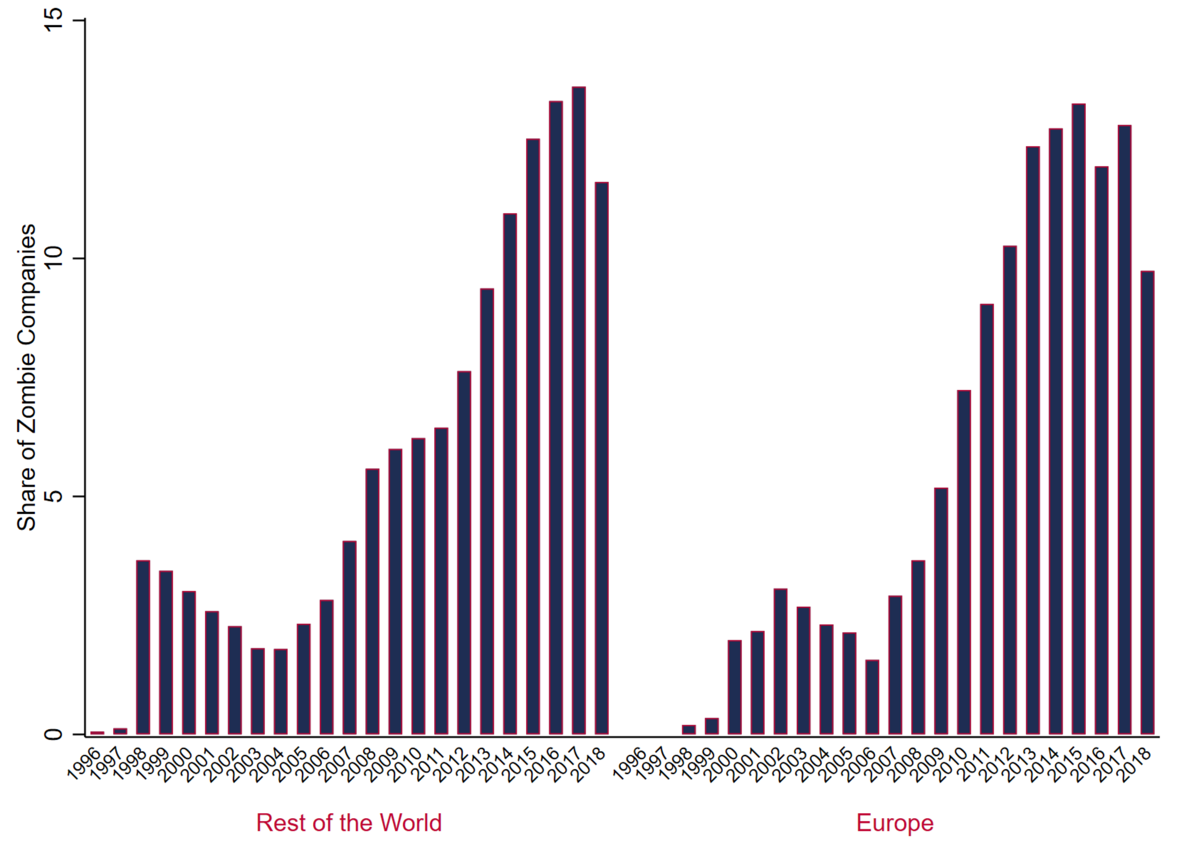
Are you a Zombie? Machine Learning Methods to Classify Unviable Firms

Angela De Martiis* Thomas Heil Franziska Peter
University of Bern, Zeppelin University

Motivation

1) Rise in *zombie* firms, alive due to bank support [1]
2) Regulatory concern, first order issue since GFC
3) Machine learning to classify, predict zombies



Contribution

1) Random Forests (RF) to classify/predict zombies
2) Examine differences/similarities between zombies and non-zombies, Europe and US, crisis/non-crisis

To do so ⇒ Large datasets of European, US firms
⇒ Machine learning methods
(Tree-based models)

Objective ⇒ Tool that can serve central banks deploy credit more efficiently avoiding misallocation

Methodology

CART algorithm [2] to find best input, split point s at each iteration

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k),$$

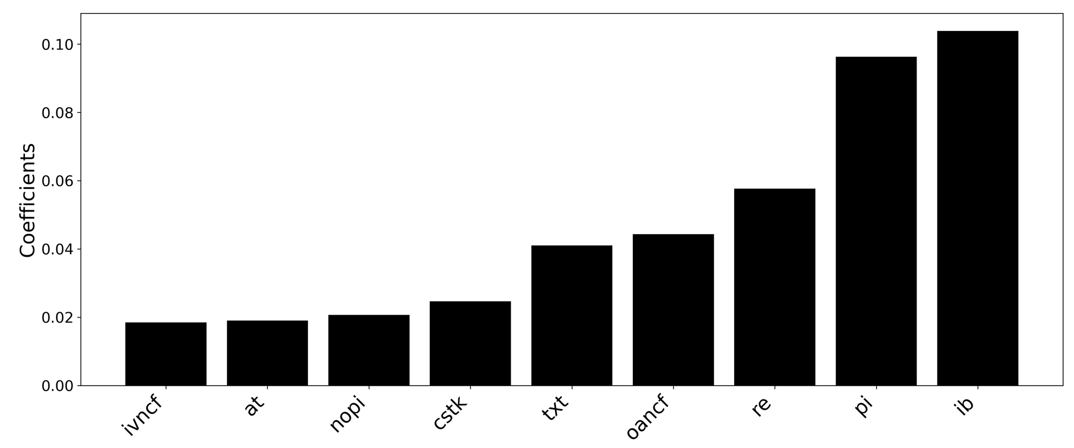
Cross-entropy as standard loss function:

$$L(p) = -\sum_{k=1}^{K} \hat{p}_{mk} log(\hat{p}_{mk}),$$

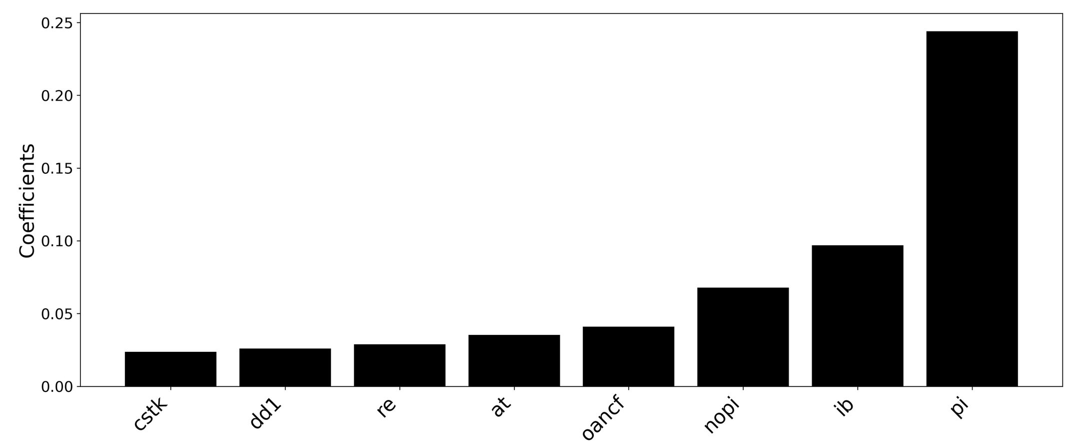
Given l split var, s split point, define pair of regions:

$$R_1(l, s) = \{X|X_l \leq s\} \; \& \; R_2(l, s) = \{X|X_l > s\}.$$

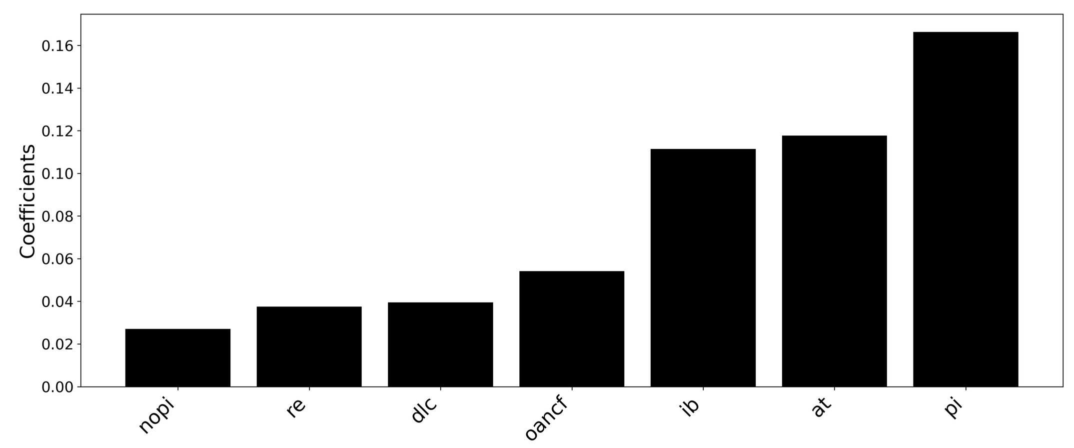
Which Firm Characteristics Matter? Results Random Forests



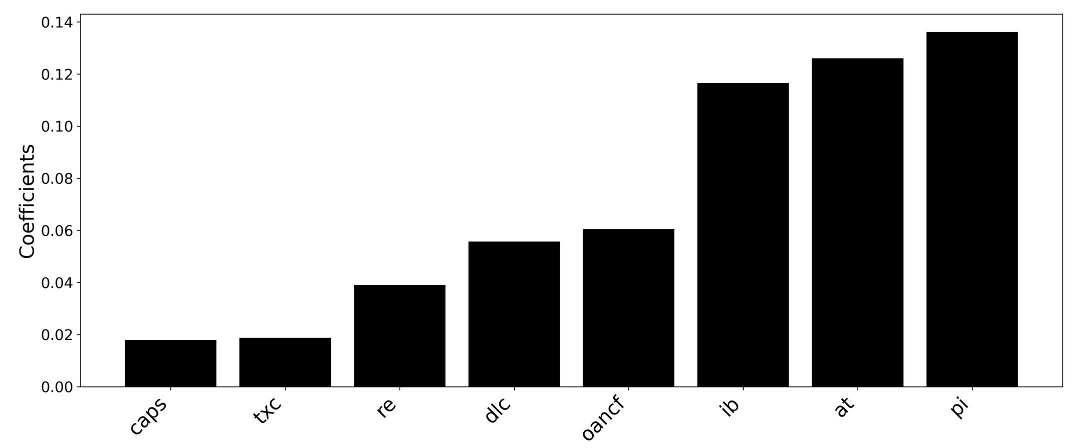
Informativeness Zombie Features Europe 2007



Informativeness Zombie Features Europe 2016



Informativeness Zombie Features US 2007



Informativeness Zombie Features US 2016

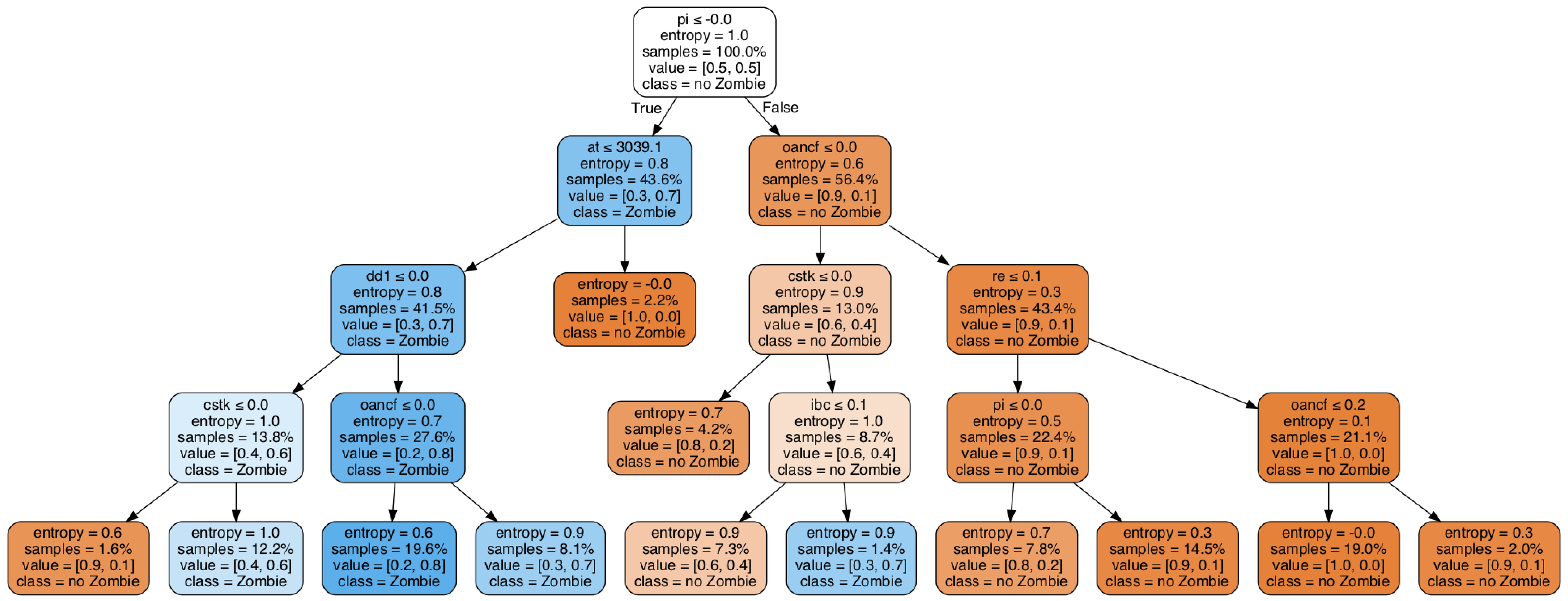
Firm characteristics that matter to predict zombie firms (higher coefficients)

⇒ Pretax income, *pi* (Europe, US, crisis/non), Operating activities, *oancf* (Europe, US, crisis/non)

⇒ Long-term debt, *dd1* (Europe, non-crisis), Short-term debt, *dlc* (US, crisis/non)

⇒ Total assets, *at* (US, crisis/non). Income-related features are the most informative (Europe, US)

Decision Tree



Example Binary Tree Europe (2016): Zombies (blue) and Non-Zombies (orange)

⇒ Pretax income (*pi*) most important split to classify zombie firms

⇒ If $x_i \leq$ *split point* (*pi*) is correct, we follow True branch, otherwise False branch

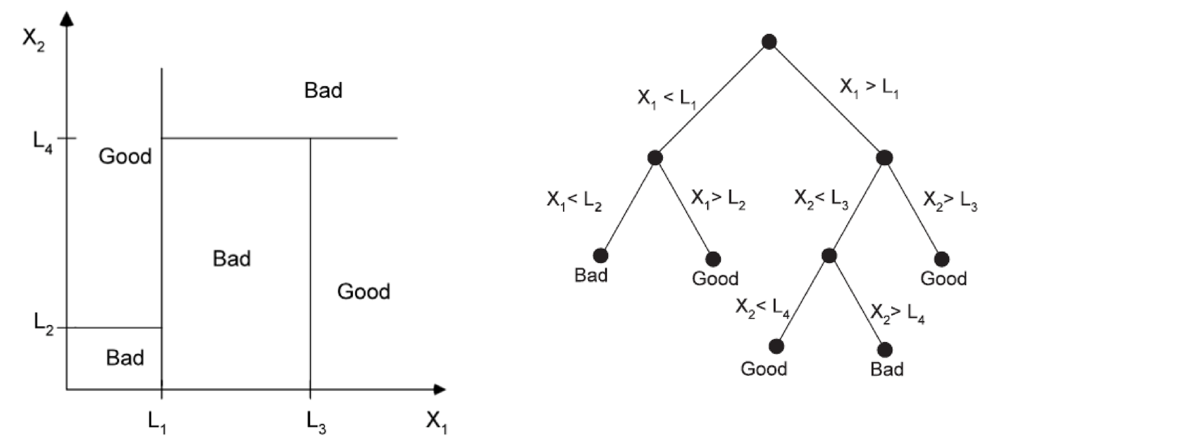
⇒ Entropy measures nodes' purity. Deeper color show how well variable separates two classes

Machine Learning Methods

Many explanatory variables can be used to predict zombie status (accounting, market data)

Standard approach ⇒ Humans perform selection
⇒ Undisciplined with many vars
⇒ Implies a priori assumptions

ML approach ⇒ Automated selection
⇒ Recursive splitting algorithm that generates trees
⇒ RF to find informative features
(RF hyperparameters: 3-fold CV)



Classification Tree Example (Kim-Khandani-Lo 2010)

	Europe			USA		
Years	Acc. (%)	Matrix		Acc. (%)	Matrix	
2007	90.89	486	0.0	90.50	502	25
		49	3.0		34	60
2016	85.60	402	31	91.61	423	22
		45	50		21	47

Prediction Results Zombie Firms (Authors' estimations)

Data and Empirical Measures

1) European and US public firms (Compustat Global/North America, Datastream)
2) 15000 obs. per year Europe, 6000 obs. per year US sample. 70 variables per company-year
3) Two cross-sections: 2007 (crisis), 2016 (healthy)
4) Zombie firms identification follows [3] and [4]

References

[1] Caballero, R. J., Hoshi, T., & Kashyap, A. K. (2008). Zombie lending and depressed restructuring in Japan. *American economic review*, 98(5), 1943-77.

[2] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

[3] Acharya, V., Crosignani, M., Eisert, T., & Eufinger, C. (2020). Zombie Credit and (Dis-)Inflation: Evidence from Europe.*National Bureau of Economic Research 27158.*

[4] Banerjee, R., & Hofmann, B. (2020). Corporate zombies: Anatomy and life cycle. *BIS Working Papers No 882.*