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:
\[ R_1(l, s) = \{X | x_i \leq s\} \text{ & } R_2(l, s) = \{X | x_i > s\}. \]

Which Firm Characteristics Matter? Results Random Forests

Firm characteristics that matter to predict zombie firms (higher coefficients)
⇒ Pretax income, \( pi \) (Europe, US, crisis/non), Operating activities, \( oanf \) (Europe, US, crisis/non)
⇒ Long-term debt, \( ddl \) (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 \( x_i \leq \text{split point (} \) is correct, we follow True branch, otherwise False branch
⇒ Entropy measures nodes’ purity. Deeper color show how well variable separates two classes

Classification Tree Example (Kim-Khandani-Lo 2010)

Classification and regression trees. CRC press.

Prediction Results Zombie Firms (Authors’ estimations)

Data and Empirical Measures
1) European and US public firms (Compustat Global, North America, Datastream)
2) 15,000 obs. per year Europe, 6,000 obs. per year US sample. 70 variables per company-year
3) Two cross-sections: 2007 (crisis), 2016 (healthy)

References

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)

Table: Classification Tree Example (Kim-Khandani-Lo 2010)

<table>
<thead>
<tr>
<th>Years</th>
<th>Europe</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>90.89</td>
<td>98.62</td>
</tr>
<tr>
<td>2016</td>
<td>85.60</td>
<td>90.95</td>
</tr>
</tbody>
</table>

Prediction Results Zombie Firms (Authors’ estimations)

<table>
<thead>
<tr>
<th>Years</th>
<th>Europe</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>90.89</td>
<td>98.62</td>
</tr>
<tr>
<td>2016</td>
<td>85.60</td>
<td>90.95</td>
</tr>
</tbody>
</table>