China's Financial Statement Fraud Detection

With Statistical Learning

Bin Pan

School of Nobody
University of Whatever

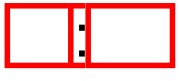
12 June 2018

Overview

Main Problem

- How best can investors, auditors, financial analysts, and regultors detect misstatements?
- The paper examine the characteristics of misstating firms along five dimensions: accrual quality, financial performance, non-financial measures, off-balance sheet activities, and market-based measures.

"I believe that machine learning (ML) will have a dramatic impact on the field of economics within a short time frame"
-Susan Athey[1]



Previous Literature

- Bell and Carcello[2] uses a sample of 77 fraud engagements and 305 nonfraud engagements to estimate the likelihood of fraudulent financial report. The significant risk factors: weak internal control environment, rapid company growth, inadequate or inconsistent relative profitability.
- Dechow and Ge[4] examine the 2190 AAERs released between 1982 to 2005 and develop a model to predict misstatements.Their model's precision is 73.8% and its accuracy is 61.7%
- Period, Bowen and[5] use a dataset with 51 fraud firms, 15,934 non-fraud firm years, and 109 explanatory variables from prior research.

Sample and Data

- Dataset from CSMAR Database[1]
- The dataset releases a list of companies who violated the laws and were fined by CSRC, SSE and SZE.
- Period: 2001-2017
- Length of the list: 5639 (After clear the data)
-

[1] CSMAR collects data from Chinese financial markets. Its counterpart in China is Wind Database

Type of Misstatements

Table: Summary of Misstatements

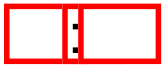
Type of Misstatement	TypeID	Freq	Percent	Cum
Others	P2599	2410	25.53%	100.00%
Delayed disclosure	P2504	1952	20.68%	33.80%
Major Omission	P2505	1415	14.99%	48.79%
Misleading statement	P2503	969	10.27%	13.12%
Illegal Stock Trading	P2512	919	9.74%	67.64%
Improper Accounting Treatment	P2515	458	4.85%	74.47%
Dishonest Information Release	P2506	333	3.53%	52.31%
Invade of company assets	P2510	271	2.87%	56.43%
Fictional Assets	P2501	230	2.44%	2.44%
Illegal Guarantee	P2514	166	1.76%	69.62%
Secret Deal	P2511	140	1.48%	57.91%
Unauthorized Change of Fund Use	P2509	104	1.10%	53.55%
False Assets	P2502	39	0.41%	2.85%
Manipulating Stock Price	P2513	20	0.21%	67.86%
Fraud Listing	P2507	8	0.08%	52.40%
Capital Violation	P2508	5	0.05%	52.45%
**	Sum	9439	100%	**

Source: CSMAR

Variables

- Y
  - $Y_{it} = 1$ if company i was caught in year t
    Otherwise $Y_{it} = 0$
- X
  - Ratios based on financial reports:
    - Accrual Quality: Measure of AQ
    - Financial performance: Growth rate etc.
  - off-balance-sheet factors:
    - Non-financial measures:
      The trend of employment rate
    - Off-balance sheet activities
      Change of Chief Officers
  - Other factors
    - Market-based measures
      PE, Tobin Q, Book to market value
-

Tools

- Combat the imbalanced dataset (data rarity-¿overfitting)
  - Use SMOTE(Synthetic Minority Oversampling Technique)[3]
  - The previous literature use undersampling method.
- Classifications
  - Logit Model
  - C4.5 (Can deal with Continuous variables)
  - Random Forest (Python *sklearn* packages)
- Evaluating Mode
  - Precision, Recall and F1
  - Confusion Matrix
  - ROC Curve
  - 

Timelines

- June 11 - June.21 finish the work
- Or be finished

Reference

[1] Susan Athey. The impact of machine learning on economics. 2018.

[2] Timothy B. Bell and Joseph V. Carcello. A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing A Journal of Practice*, 19(1):169–184, 2000.

[3] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, 2002.

[4] Patricia M. Dechow, G. E. Weili, Chad R. Larson, and Richard G. Sloan. Predicting material accounting misstatements. *Contemporary Accounting Research*, 28(1):17–82, 2011.

[5] Johan Perols, Robert M Bowen, Carsten Zimmermann, and Basamba Samba. Finding needles in a haystack: Using data analytics to improve fraud prediction. *Social Science Electronic Publishing*, 2015.