Redoing ratio analysis with clustering techniques

Kexing Ding

Xuan Peng

Miklos Vasarhelyi

Yunsen Wang

Rutgers University-Newark

Peer selection with clustering

- We apply a machine learning technique (clustering) on financial ratios to identify peer firm.
  - SIC/ NAICS codes are commonly used to identify peers
    - Obsolete
    - Infrequent update
    - Focus on production process
  - Ratio selection: choose the set of financial ratios based on research objective
    - For example, if the purpose of the research is to predict a specific corporate event, the selected financial ratios should be related to this event.
  - We do two tests: (1) detecting material accounting misstatements and (2) predicting corporate bankruptcies

1. Detecting material accounting misstatements

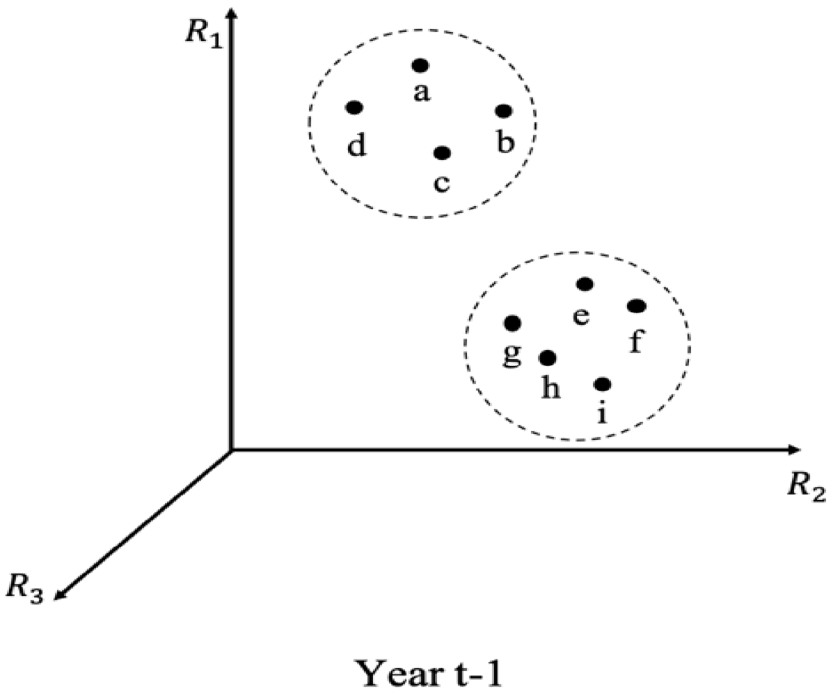
– Ratios in clustering analysis: 8 ratios that are associated with accounting misstatements

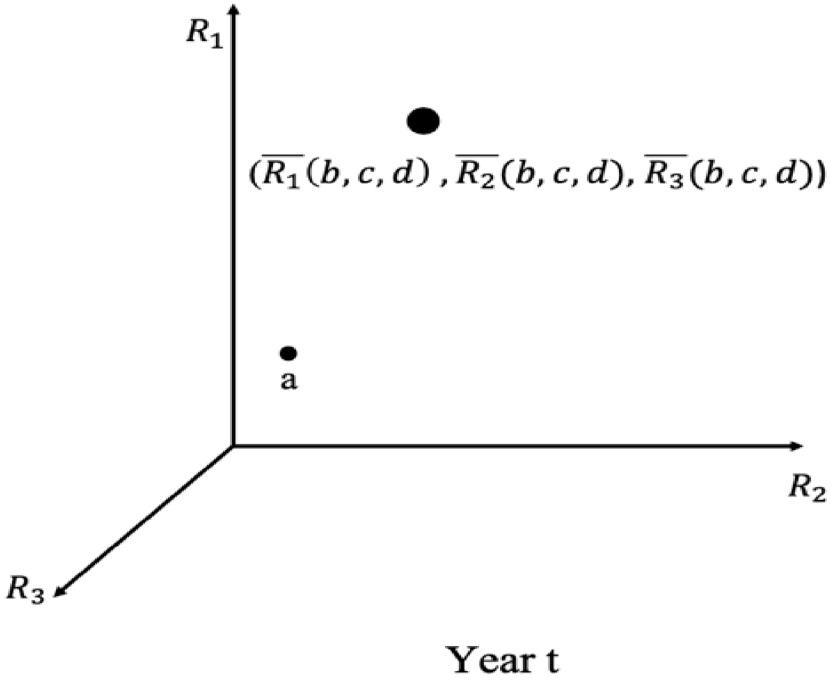
- Sales in receivables (Receivables/Sales),
- Gross Margin (Sales-Costs of Goods Sold/Sales),
- Asset Quality (1 - (Current Assets + PPE)/Total Assets),
- Sales Growth ($Sales_t/Sales_{t-1}$),
- Depreciation Rate (Depreciation/NetPPE),
- SGA Rate (Sales, general, and administrative expenses/Sales),
- Leverage (Total Debt/Total Assets),
- Accruals (Total Accruals/ Total Assets).

1. Detecting material accounting misstatements

– Clustering

- The clustering algorithm groups firms that are similar to each other in year $t-1$.
- If in year $t,$ a firm reports differently from the average of its peers, we "redflag" it, and construct a variable, DevScore, to capture its difference from its peers in year t.





1. Detecting material accounting misstatements

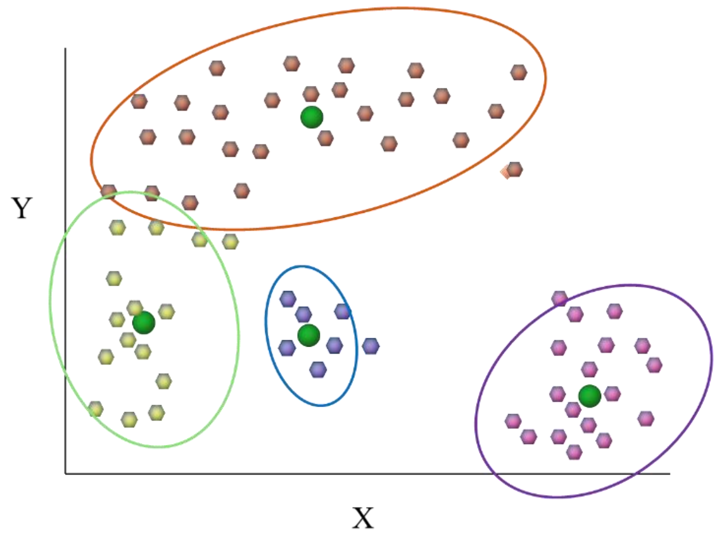
– Out-of-sample test results

- The proportion of actual misstatements in decile groups that are ranked by the estimated misstatement probability by each model.
- Results suggest that the model with clustering DevScore can detect misstatements most efficiently: 40.70% misstating firms in 10th decile; around 85% of misstating firms in the top 5 deciles

Decile	Dechow model (DM)	Dechow model with clustering DevScore	Dechow model with SIC DevScore	Dechow model with NAICS DevScore
1-5	17.44%	**15.12%**	19.76%	20.94%
6	10.47%	**4.65%**	6.98%	4.65%
7	11.63%	**9.30%**	10.47%	12.79%
8	12.79%	**12.79%**	11.63%	13.95%
9	20.93%	**17.44%**	25.58%	20.93%
10	26.74%	**40.70%**	25.58%	26.74%
Total	86 misstatements			

2. Predicting bankruptcy

- Ratios in clustering analysis: 5 ratios that are associated with corporate bankruptcy
  - Net working capital to total assets ratio (WC/TA)
  - Retained earnings to total assets ratio (RE/TA)
  - Earnings before interest and taxes to total assets ratio (EBIT/TA)
  - Market value of equity to book value of total liabilities (ME/BL)
  - Sales to total assets (SALE/TA)
- Clustering: peers with similar "bankruptcy" ratios
- Compute *DevScore to capture a firm's differences from its peers*



1. Predicting bankruptcy

– Out-of-sample test results

– The proportion of actual bankruptcies in decile groups that are ranked by the estimated bankruptcy probability by each model.

– Results suggest that the Shumway model with clustering DevScore outperforms other models in bankruptcy predicting: 88% of bankruptcy firms in the top decile, higher than other three models.

Decile	Shumway model	Shumway model with clustering DevScore	Shumway model with SIC DevScore	Shumway model with NAICS DevScore
1-5	2.89%	**0.97%**	1.93%	1.93%
6	0.96%	**0.00%**	0.96%	2.88%
7	0.00%	**0.96%**	2.88%	0.96%
8	2.88%	**1.92%**	2.88%	1.92%
9	11.54%	**7.69%**	10.58%	12.50%
10	81.73%	**88.46%**	80.77%	79.81%
Total	104 bankruptcies			

Thank you!