Research Question
000

Research Design
000

Results
00000

Ideas
0

Appendix
00

Unearthing Financial Statement Fraud: Insights from News Coverage Analysis

Jianqing Fan   Qingfu Liu   Bo Wang   Kaixin Zhengd
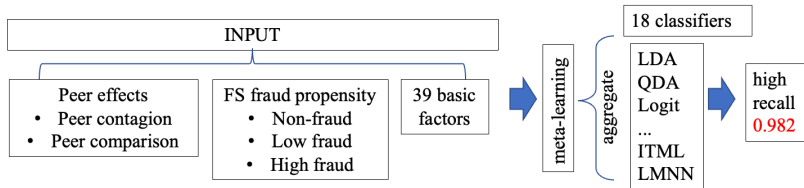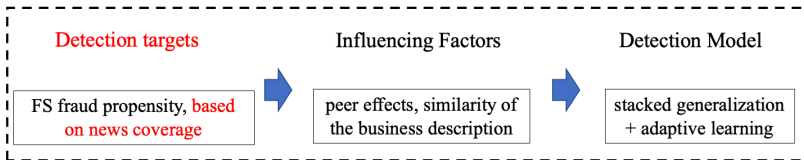
Management Science, 2025

解读人：冯丽璇

武汉大学金融系

October 12, 2025

Research Question
ooo

Research Design
ooo

Results
ooooo

Ideas
o

Appendix
oo

Summary

Unearthing Financial Statement Fraud: Insights from News Coverage Analysis

The Detection Process of Financial Statement Fraud

Detection targets

FS fraud propensity, based on news coverage

Influencing Factors

peer effects, similarity of the business description

Detection Model

stacked generalization + adaptive learning

INPUT

Peer effects
• Peer contagion
• Peer comparison

FS fraud propensity
• Non-fraud
• Low fraud
• High fraud

39 basic factors

meta-learning

aggregate

18 classifiers

LDA
QDA
Logit
...
ITML
LMNN

high recall
0.982

Research Question

1. How can financial statement fraud be better defined?
   - considering the role of news coverage
2. Does Peer effects help to detect FS fraud?
   - yes, use similarity of the business description to define peer effects.
3. How can a more effective financial fraud prediction model be constructed?
   - utilizes two top-level machine learning algorithms, stacked generalization and adaptive learning.

Motivation

- Financial statement (FS) fraud damages the interests of investors and the credit cornerstone of the capital market, studying the detection of FS fraud is essential.
- Measurement of detection targets
    - Prior research have selection bias problem: **FS fraud events** that have already occurred, still in the stage of investigation are often classified as nonfraud samples.
- This study construct a new measure of **FS fraud behavior** based on news coverage.
- Peer effects are very common in financial misconduct, yet current research still has limitations in grouping firms.
- The application of machine learning algorithms is limited to classic methods.

Research Question
○○●

Research Design
○○○

Results
○○○○○

Ideas
○

Appendix
○○

Contribution

1. Contributes to the research on the samples or measurement of FS fraud
   - Prior: companies that have financial restatements, are subject to penalties, or receive nonstandard audit opinions.
   - Extend: FSFP method, built based on the number of news reports, its nonzero value covers more potentially fraudulent samples.
2. Contributes to the method for grouping peers.
   - Prior: classify listed companies into groups based on the same industry or the same region
   - Extend: based on the operating business in the annual reports.
3. Contributes to the detection method of FS fraud.
   - Prior: often only use the financial ratio of the listed company as an input variable.
   - Extend: develop a fraud risk index that leverages news coverage for the detection of FS fraud, and integrate news-based metrics to formulate peer contagion factors.

Research Question
000

Research Design
●○○

Results
○○○○○

Ideas
○

Appendix
○○

Research Hypothesis

1. New Measures of FS Fraud Based on News Coverage.

   - Ambiguity in the definition of FS fraud, current measures reflect FS fraud events instead of behavior.
   - News coverage can alleviate the FS fraud caused by information asymmetry by exposing these irregular behaviors.

2. Peer Effects and FS Fraud Risk Factors.

   - Defines peers as a group of firms with a high similarity in their business operations.
   - FS fraud behavior tends to spread among firms with similar businesses.

3. Detection Model of FS Fraud

   - The SG-AL algorithm integrates stacked generalization and adaptive learning, leveraging diverse base classifiers while reducing information lag in dynamic environments.

Research Question
000

Research Design
0●0

Results
00000

Ideas
0

Appendix
00

Measurement of FSFP

- China Economic News Database (CEND), from January 1, 2001, to December 31, 2022.

$$FSFP_{it} = \frac{1}{N_{it}} \sum_{j=1}^{M_{it}} (ATF \times IIF)_{itj},$$

  - M_{it} is the total number of FS fraud–related news reports, and N_{it} is the number of all news reports of listed company i in period t, whether it is FS fraud–related or not.
  - ATF and IIF: $ATF = \sum_{k \in K} TF_k$, $IIF = \frac{1}{\#firm_{tj}/\#firm_t}$
    ATF is the total frequency of terms in set K in an article, and IIF gives lower weight to articles covering many firms.

Research Question
000

Research Design
00●

Results
00000

Ideas
0

Appendix
00

Grouping Listed Companies

- Current grouping method has many shortcomings
  - industry classification cannot fully reflect the operations of listed companies.
  - region cannot adapt to the enterprise organization mode and market environment
  - shareholders can only reflect a small aspect of the relatedness of listed companies
- the text of the annual report reflect the dynamic and continuous changes in the linkage among listed companies.
- Cosine similarity of business texts, CNM algorithm to identify groups.
  - Peer Contagion Factors: the influence of peer companies' fraudulent behaviors on the target company.
  - Peer Comparison Factors: the financial and operational performance of a company relative to its peer group

# Result 1: Effectiveness Test of FSFP Measure

- FSFP shows predictive power, with significant positive coefficients for leading values of dummy measures.

- If the measure of FS fraud reflects well FS fraud behaviors, then the measure is negatively correlated with regulatory strictness.

Dependent variable	$logit(Penalty_{i,t})$ (1)	$logit(Restate_{i,t})$ (2)	$logit(NsAudit_{i,t})$ (3)
$FSFP_{i,t-1}$	0.112***	0.054**	0.049***
	(4.179)	(2.436)	(2.826)
$FSFP_{i,t-2}$	0.039***	0.021**	0.095***
	(2.878)	(1.966)	(12.537)
$FSFP_{i,t-3}$	0.028**	0.005	0.047***
	(2.314)	(1.554)	(2.997)
$FSFP_{i,t-4}$	0.027	−0.024	0.016
	(1.750)	(−1.552)	(0.498)
$FSFP_{i,t-5}$	0.005	−0.003	−0.001
	(0.411)	(−0.925)	(−0.653)
Constant	0.009	−0.114	−0.022
	(1.013)	(−1.085)	(−1.514)
Industry fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓
Observations	49,779	49,779	49,779
Adjusted R^2	0.248	0.286	0.337

Result 1: Effectiveness Test of FSFP Measure

- The FSFP measure effectively captures FS fraud events, with high values corresponding to confirmed fraud cases.

- 000921.SZ: 海信家电

- 600477.SH: 杭萧钢构

- 002200.SZ: 科华生物

- 300268.SZ: 万福生科

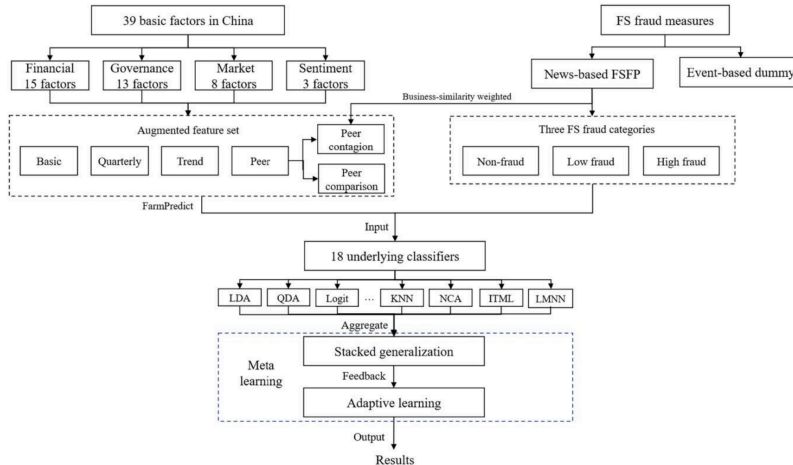- 002450.SZ: 康得新材

- 600518.SH: 康美药业

- 002383.SZ: 合众思壮
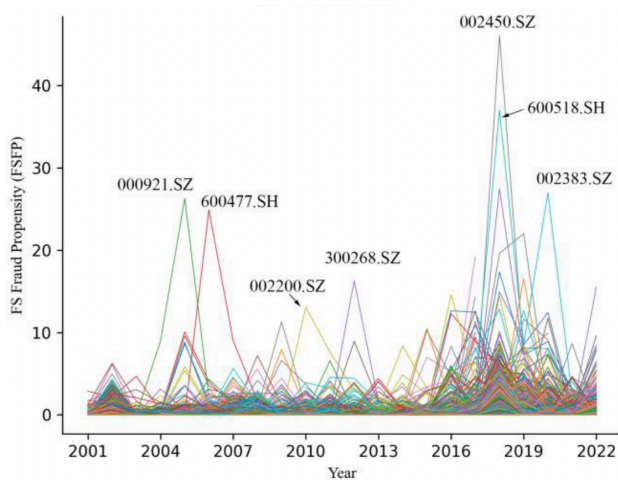
Result 2: Grouping Listed Companies

- Peer effects are more pronounced in high-competition groups, supporting the hypothesis that market competition drives the contagion of FS fraud.

Dependent variable	$ln(1 + FSFP)$ (1)	$logit(Penalty)$ (2)	$ln(1 + FSFP)$ (3)	$ln(1 + FSFP)$ (4)
Panel A: Mechanism examination with high-competition subsamples				
$ln(1 + \overline{FSFP})$	0.089***			
	(3.138)			
$\overline{Penalty}$		0.175***		
		(5.136)		
$ln(1 + \overline{FSFP})_{Similarity\,weighted}$			0.201***	
			(5.320)	
$ln(1 + \overline{FSFP})_{CSRC\,industry}$				0.053***
				(2.832)
Constant	0.094	0.121	0.513	0.165**
	(0.769)	(1.143)	(1.245)	(2.195)
Firm fixed effects	✓	✓	✓	✓
Group-year fixed effects	✓	✓	✓	✓
Observations	56,213	56,213	56,213	56,213
Adjusted R^2	0.074	0.078	0.195	0.095
Panel B: Mechanism examination with low-competition subsamples				
$ln(1 + \overline{FSFP})$	0.029*			
	(1.810)			
$\overline{Penalty}$		0.023***		
		(2.789)		
$ln(1 + \overline{FSFP})_{Similarity\,weighted}$			0.074**	
			(2.023)	
$ln(1 + \overline{FSFP})_{CSRC\,industry}$				0.012**
				(1.758)
Constant	0.091	0.180**	0.834**	0.306***
	(1.110)	(2.263)	(2.351)	(2.740)
Firm fixed effects	✓	✓	✓	✓
Group-year fixed effects	✓	✓	✓	✓
Observations	56,213	56,213	56,213	56,213
Adjusted R^2	0.088	0.114	0.172	0.082

Research Question
000

Research Design
000

Results
00000

Ideas
0

Appendix
00

Result 3: Detection Model of FS Fraud

Research Question
000

Research Design
000

Results
0000●

Ideas
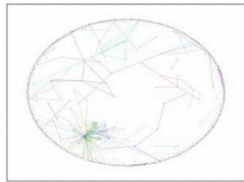0

Appendix
00

Detection Model of FS Fraud

- The overall recall of regulatory penalty–based fraud is $0.982 = 1{,}266/(1{,}266 + 23)$.

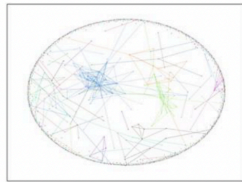|			Panel A: Numbers of true and false predictions		
Year	Number of observations	Number of true frauds	Number of false frauds	Number of false nonfrauds	Number of true nonfrauds
2006	1,406	23	22	0	1,361
2007	1,511	24	17	1	1,469
2008	1,565	19	10	0	1,536
2009	1,712	57	27	1	1,627
2010	2,061	39	35	0	1,987
2011	2,291	41	35	0	2,215
2012	2,420	65	54	1	2,300
2013	2,462	95	85	1	2,281
2014	2,576	77	72	1	2,426
2015	2,763	98	92	4	2,569
2016	3,185	86	52	3	3,044
2017	3,559	98	48	1	3,412
2018	3,644	150	74	4	3,416
2019	3,841	134	87	0	3,620
2020	4,280	98	82	2	4,098
2021	4,697	86	77	2	4,532
2022	4,997	76	67	2	4,852
Total	48,970	1,266	936	23	46,745

Ideas

- 卖空报告或机构预测作为标签 Y（恒大暴雷之前，美国空头机构香橼在 2012 年发布了做空报告）
  - 若专业卖空机构（如 Muddy Waters, Hindenburg）发布了质疑报告，可作为"市场预判造假"的 proxy
  - 即使尚未被证实，也可以作为"潜在欺诈怀疑"标签
- 识别同行企业
  - AI 分析企业"主营业务"报告文本划分同行企业
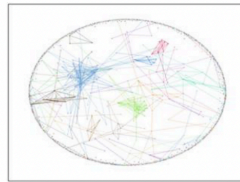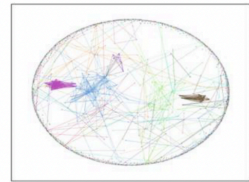 - 企业专利文本（标题、摘要）的语义相似性，识别技术上相似、竞争或合作关系密切的同行企业群。

Research Question
000

Research Design
000

Results
00000

Ideas
0

Appendix
●0

Visualization of Business Networks

Research Question
000

Research Design
000

Results
00000

Ideas
0

Appendix
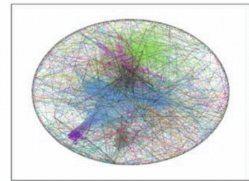0●
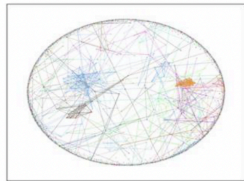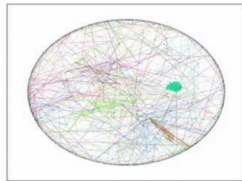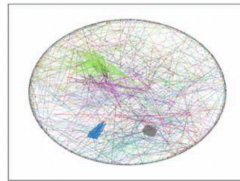
同伴影响因素计算

- 公司 A 的财务报表欺诈倾向性（FSFP）= 0.1
- 公司 B 的财务报表欺诈倾向性（FSFP）= 0.2
- 公司 C 的财务报表欺诈倾向性（FSFP）= 0.3
- 公司 A 与公司 B 的业务相似度 = 0.8
- 公司 A 与公司 C 的业务相似度 = 0.6
- 公司 A
  - 同伴传染因素 $= 0.8 \times 0.2 + 0.6 \times 0.3 = 0.26$
  - 同伴组的平均 FSFP $= \frac{0.2+0.3}{2} = 0.25$
  - 同伴比较因素 $= 0.1 - 0.25 = -0.15$