PAC ANNUAL CONFERENCE

PROFESSIONAL ACCOUNTING FUTURES

December 1, 2023

Perspectives on Fraud Prediction Research: From Firms' Misreporting to Misinformation in the Macroeconomy

Daniel Beneish,

Professor, Indiana University

Professional Accounting Centre

UNIVERSITY OF TORONTO

MISSISSAUGA

Agenda

1. Introduction to the M-Score
2. Some Important Detections
3. Some Important Misses
4. Comparing Fraud Prediction Models
5. From Micro to Macro: the likelihood of misreporting in the economy

Days Sales in Receivables (DSR)

Day Sales in Receivables (DSR)	$\frac{\text{Average Receivables}}{\text{Net Sales/365}}$	Shows how many days it takes to collect sales

Interpret these data: Year$_t$ Year$_{t-1}$

DSR	30.4	30.6

DSR	42.4	30.6

If $DSRI=DSR_T/DSR_{T-1}$, What does a value = 1 indicate?
What does a value > 1 indicate?

Gross Margin Percentage

$$\text{Gross Margin Percentage} = \frac{(\text{Net Sales} - \text{Cost of Goods Sold})}{\text{Net Sales}}$$

Gross Margin Percentage is the mark-up over cost at which the firm can sell.

What does the Gross Margin Percentage really mean?

	WMT	Lululemon
GM%	25%	55%

If $GMI = GM_{T-1}/GM_T$, What does a value > 1 indicate?

	2023	2022
GM%	29.4%	33.1%

If GM in any year is negative, $GMI = 1 + (GM_{T-1} - GM_T)$.

Leverage

Liabilities to Assets Ratio	$\frac{\text{Total } Liabilities}{\text{Total } Assets}$	Indicates the percentage of assets financed through borrowing. If we denote the Debt-to-Equity Ratio as X, this ratio equals X/(1+X).

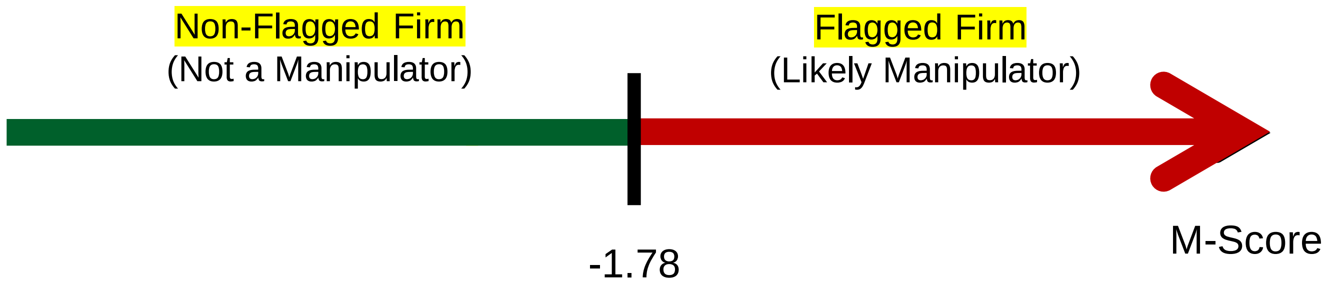
Interpret

	2023	2022
LEV%	69.1%	52.1%

If $LEVI=LEV_T/LEV_{T-1}$, What does a value > 1 indicate?

Earnings Manipulation

M – Score = -4.840 + .920 Days is Receivables index
+ .528 Gross Margin Index
+ .404 Asset quality index
+ .892 Sales growth
+ .115 Depreciation index
- .172 SGA index
+ 4.679 Accruals to total assets
- .327 Leverage index



M.D. Beneish-M-Score Presentation 2023

Using the Calculator—Lucent Technologies
https://apps.kelley.iu.edu/Beneish/MScore/MScoreInput

2. The data can be input using any currency or units (thousands, millions) so long as currency and units are kept the same for all input numbers.
3. The calculator computes the M-Score, the odds ratio that the firm is a manipulator, and provides an assessment based on whether the model's variables are out of line relative to sample averages. The M-Score is calculated as in Beneish, Lee Nichols 2013, *Financial Analysts Journal*.
4. The use of the calculator is illustrated with data from Sunbeam and Tesla; the input and output are reproduced in the Sunbeam example. The spreadsheet that you can complete is in the Telsa file.
5. Using the example in the Tesla spreadsheet, you can copy the 24 numeric values and paste them into any field in the table below.

Financial Statement Inputs	Lucent		
	1999	1998	Acceptable Input
Acc. Receivable (Trade), Net	10438	7405	>0
Current Assets	21931	15784	>0
Current Liabilities	11778	10885	>0
Total Assets	38775	29363	>0
PPE Net	6847	5693	>0
Long Term Debt	4162	2409	>=0
Sales (Net)	38303	31806	>0
Depreciation Expense	1806	1411	>0
Cost of Goods Sold	19688	16715	>0
SGA Expense	8417	6867	>0
CFO	421	1366	Any number
Net Income (excl. Extr. Items)	3833	2287	Any number

Submit

Lucent FY (September 1999) M-Score

Beneish M-Score Calculator

Perform New Calculation

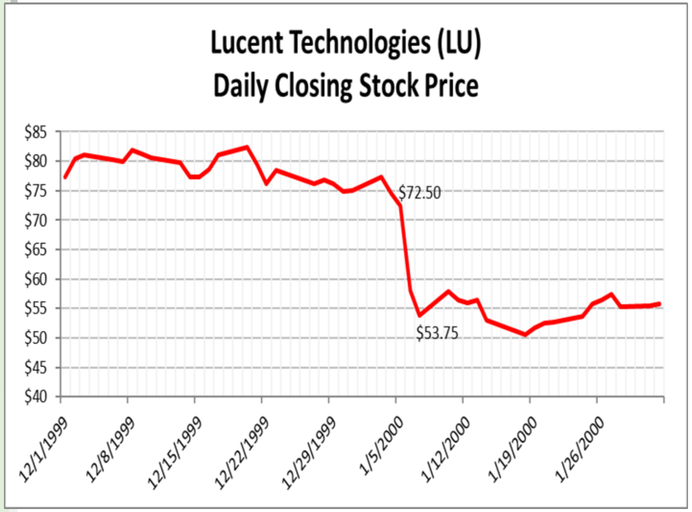
Lucent

Analysis	Variable	Assessment
Day Sales in Receivable Index	1.170	Assess revenue recognition
Gross Margin Index	0.976	Neutral
Asset Quality Index	0.960	Neutral
Sales Growth Index	1.204	High sales growth
Depreciation Index	0.952	Neutral
SG&A Index	1.018	Neutral
Accruals to Total Assets	0.088	Assess changes in working capital
Leverage Index	0.908	Neutral

M-Score	-1.736	Likely Manipulator
Estimated Probability	0.041	

Odds Ratio	5.98 to 1

Additional Computations	1999	1998
Days in receivables	98.10	83.81
Gross Margin	48.6%	47.4%
Depreciation Rate	20.9%	19.9%
Capital Intensity (PPE/A)	17.7%	19.4%
SG&A to Sales	22.0%	21.6%
Leverage ((CL+LTD/A))	41.1%	45.3%
Accruals to total assets	8.8%	3.1%
Profit Margin (NI/S)	10.0%	7.2%
Asset Turnover (S/A)	0.99	1.08
Return on Assets (NI/A)	9.9%	7.8%
Return on Equity (NI/SE)	16.8%	14.2%



Satyam Computer Services—India
Firm loses 94 % of its value over January 7 to 9, 2009

From AP Wires @ CNBC https://www.cnbc.com/id/28567215

Shares in India's Satyam Computer Services slumped more than 70 percent Friday (Jan 9 2009) and dragged the 30-share Bombay Stock Exchange index down more than 2 percent.

Satyam, which had plunged nearly 80 percent on Wednesday (Jan 7 2009) after the outsourcer said it falsely inflated profits for many years, dropped as much as 71.2 percent to 11.50 rupees when trading resumed after a holiday on Thursday.

Satyam	*2008*	2007	2006
Day Sales in Receivables Index	1.037	1.147	1.054
Gross Margin Index	1.058	1.074	0.491
Asset Quality Index	1.546	7.042	0.039
Sales Growth Index	1.311	1.301	1.421
Depreciation Index	1.349	1.293	1.167
SG&A Index	0.951	0.992	0.909
Accruals to Total Assets	0.016	0.059	0.305
Leverage Index	1.128	1.038	1.312
Model Score	-1.835	0.703	-1.354
Model Estimated Probability	0.033	0.759	0.088
Odds Ratio	4.8	110.0	12.7

Kangmei Pharmaceuticals 2017--China

Prices: Sept 1 2018 $21.88; Oct 1 2018 $12.35

Kangmei Pharma 2017	Variable	Assessment
Day Sales in Receivables Index	1.986	Assess revenue recognition
Gross Margin Index	1.169	Why are margins deteriorating?
Asset Quality Index	1.285	Assess expense capitalization
Sales Growth Index	0.812	Neutral
Depreciation Index	1.468	Declining depreciation rate
SG&A Index	1.602	Increasing expenses
Accruals to Total Assets	0.108	Assess changes in working capital
Leverage Index	1.087	Increased borrowing
M-Score	-1.109	Odds Ratio 19.37 to 1
Estimated Probability	0.134	

M-Score Calculator:
https://apps.kelley.iu.edu/Beneish/MScore/MScoreInput

Wirecard--Germany

Prices: June 17 2020 $113; June 24 2020 $15

Wirecard 2017	Variable	Assessment
Day Sales in Receivables Index	1.00	Neutral
Gross Margin Index	1.03	Neutral
Asset Quality Index	1.03	Neutral
Sales Growth Index	1.45	High Growth
Depreciation Index	0.98	Neutral
SG&A Index	1.13	Increasing expenses
Accruals to Total Assets	0.20	Assess changes in working capital—possible 'asset bloating'
Leverage Index	1.11	Increased borrowing
M-Score	--1.161	
Estimated Probability	0.123	

Odds Ratio 17.8 to 1

But not Worldcom in 2001 …

Worldcom	2001
Day Sales in Receivables Index	0.865
Gross Margin Index	1.040
Asset Quality Index	0.886
Sales Growth Index	0.900
Depreciation Index	0.876
SG&A Index	1.158
Accruals to Total Assets	0.025
Leverage Index	1.042
Model Score	-2.654
Model Estimated Probability	0.004
Odds Ratio	0.580

Or Carillion in 2016…

Carillion		
Analysis	**Variable**	**Assessment**
Day Sales in Receivable Index	1.204	Assess revenue recognition
Gross Margin Index	1.081	Why are margins deteriorating?
Asset Quality Index	0.92	Neutral
Sales Growth Index	1.112	Neutral
Depreciation Index	1.023	Neutral
SG&A Index	1.003	Neutral
Accruals to Total Assets	0.017	Neutral
Leverage Index	1.035	Neutral
M-Score	-2.113	No manipulation
Estimated Probability	0.017	
Odds Ratio	2.51 to 1	

Additional Computations	2016	2015
Days in receivables	130.53	108.41
Gross Margin	8.00%	8.60%
Depreciation Rate	23.80%	24.40%
Capital Intensity (PPE/A)	3.30%	3.60%
SG&A to Sales	5.00%	4.90%
Leverage ((CL+LTD/A))	63.40%	61.20%
Accruals to total assets	1.70%	2.00%
Profit Margin (NI/S)	3.30%	3.80%
Asset Turnover (S/A)	0.99	1.02
Return on Assets (NI/A)	3.30%	3.90%
Return on Equity (NI/SE)	9.00%	10.00%

Fraud prediction models

1. Beneish 1997-*Journal of Accounting and Public Policy*
2. Beneish 1999-*Financial Analysts' Journal*
3. Cecchini et al. 2010-*Management Science*
4. Dechow et al. 2011-*Contemporary Accounting Research*
5. Amiram et al. 2015-*Review of Accounting Studies*
6. Bao et al. *2020-Journal of Accounting Research*
7. *Alawahdi et al. 2020-WP*
8. *Chakrabarty et al. 2020-WP*
9. *Text and data models, bag of words, unexpected audit fees,*

and … random forests, support vector machines, KNN classifiers….

Genesis

Excerpt from letter to the Editors of
The Accounting Review

Re**:** "Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction"

….

This article benchmarks its work to the models in Cecchini et al. (2010, *Management Science*) and Dechow et al. (2010, *CAR*) **which they label in their abstract as the 'best current techniques' to detect fraud**. The authors use the area under the Receiver Operating Characteristic Curves (ROC) as a means of comparing models.

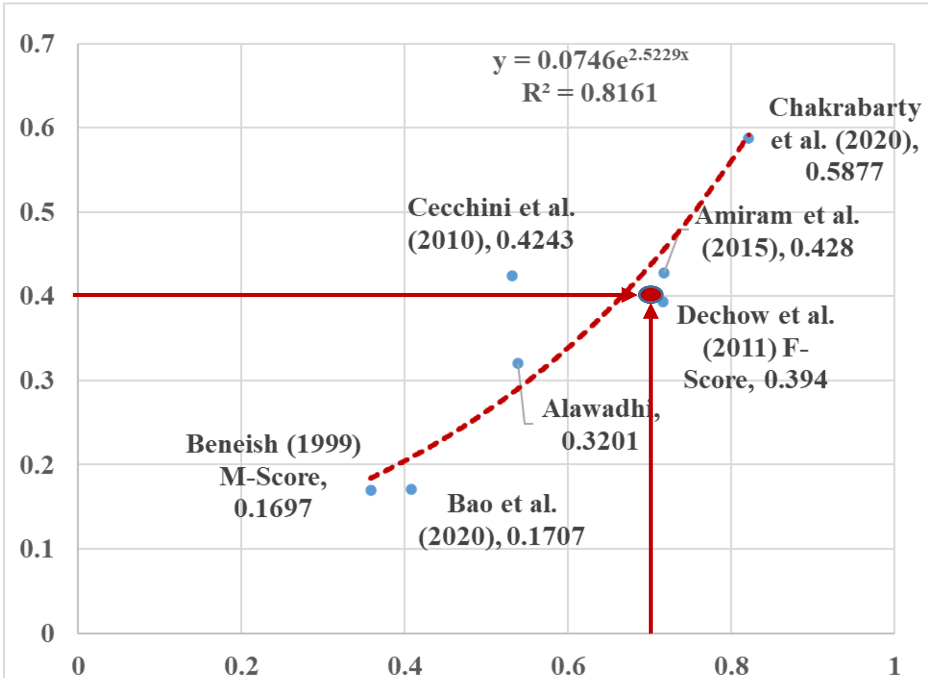
This is stunning to me because neither of these 'best' techniques has a credible record of out-of-sample prediction—I am not just talking about holdout samples I am talking about out-of-sample performance after publication. …. Results in "Finding Needles…" regarding Cecchini et al. do not help in that regard as their smaller sample has a nearly identical data period (1998-2005) to that of Cecchini et al.'s (1999-2006).

Dechow, et al. (2010, CAR) show their F-Score identifies …their overall correct classification percentage is 64%. I do not know whether the Dechow et al. model overfits post 2010, but I know that false positives are the bane of any fraud detection technology from an audit firm standpoint.

In contrast, my M-Score has accumulated a very good 20-year record of out-of-sample prediction…would yield approximately an 83% correct classification percentage. ….From a practical standpoint, the paper is used by several accounting and investing firms, and the M-Score model is part to the CFE and CFA curriculum…

False v. True Positive Rates Across Seven Models (1979-2016)

False Positive Rates



True Positive Rates

About 310 unique fraud firms so that at 70% we get ≈ 217 true positives

About 136,000 non-fraud observations so that at 40% we get ≈ 54,400 false positives

So, at (70%,40%), about 54,617 firms are flagged.
How do you tell the 217 apart?

Assessing Economic Viability—Beneish and Vorst (2022)-TAR

- Beneish and Vorst (2022) study economic viability as a trade off between benefits and costs of implementing these prediction models
- Most often what prior studies have traded off are *assumptions* of benefits and costs implied by traditional metrics:
  - AUC implemented assuming **cost equality across types of errors**
  - AUC shown to overestimate model predictive performance in studies of unbalanced samples.
  - Expected costs of misclassification (ECM)— assume **all errors of one type have the same cost** (e.g., Enron has the same cost as Vivendi, Xerox, AIPC…)
- Study adds to the literature because little is known about the cost of prediction errors:
  - the costs of false positives have not been previously studied, and
  - the costs differ for auditors, investors, and regulators.

Figure 1: Overview of the Costs & Benefits of Fraud Prediction Models

		Actual	
		Fraud	Non-Fraud
Predicted	Fraud	True Positive (Benefit is the cost of false negative that is avoided)	False Positive (Cost of false positive is incurred—Type I Error)
	Non-Fraud	False Negative (False negative cost is incurred whether model is used or not—Type II Error)	True Negative (No cost incurred whether model is used or not)

Precision	True positive /Predicted Fraud
Sensitivity or recall	True positive / Total Fraud

The net benefit or net cost of using a fraud prediction model is based on the *upper row* of the matrix where the model either predicts fraud correctly (true positive) or incorrectly (false positive). If fraud is not predicted, there is no difference in costs relative to not using a model.

Auditor benefit from identifying true positives:
Mean $156 M
Median $6.5 M

Auditor cost of identifying false positives:
Mean $1.2 M
Median $0.3 M

Auditor aggregate benefit:
$156M*217=$33.85 B
Auditor aggregate cost:
$1.2 M*54400=$65.28B

Usefulness to Auditors

- Researchers in the field invariably state that the model they propose is useful to auditors and regulators and, in some cases, also to analysts and investors.
- But, at least in the case of auditors, this is not what we observe: All models too costly for auditors to implement even in extreme subsamples in which a priori the likelihood of misreporting is greater.
- Starting in 1999, my discussions with auditors and general counsel at Andersen and two other large accounting firms revealed that litigation concerns relating to false positives created an unwillingness to use fraud models in practice unless the number of false positives could be lowered.
- My efforts back then to improve the M-Score failed because I could not increase the model's success rate without increasing the number of false positives.

Usefulness to Investors and Regulators, Researchers

- For investors, the M-Score and, when used at higher cut-offs the F-Score, are the only models providing a net benefit.
- For regulators, several models are economically viable because false positive costs are limited by the number of investigations regulators can initiate
- Lowering false positive rate is a must to achieve economic viability-- Overfitting is easy, lowering false positives is not.
- False positive rates in the range of 40 to 60 percent are many times larger than the actual incidence of misreporting in the population. Using these models as proxies leads to falsely rejecting the null hypothesis of no misreporting more frequently than warranted.
- Machine learning and AI methods are likely to improve predictions in the future **but keep on eye on identified discriminators and on the number of false positives…**

Cecchini et al. 2010—*Management Science*

Table 8 Top Five Features of SVM-FK

Feature	Weight (absolute value)	Ratio	Year	Correlation with fraud
1	0.403	*Sales/Preferred Stock, Carrying Value*	$t-1$	Positive
2	0.384	*Selling, General, and Administrative Expenses/Investments and Advances, Other*	t	Negative
3	0.275	*Total Assets/Investments and Advances, Other*	$t-1$	Positive
4	0.273	*Sales/Investments and Advances, Other*	$t-1$	Negative
5	0.245	*Total Assets/Short-Term Investments*	t	Positive

Purda and Skillicorn 2015 —*CAR*

Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection

Appendix 2

Rank-ordered list of most predictive words

acquisitions	taxes	from
acquisition	capital	facility
months	revenue	business
sales	under	cash
legal	expenses	income
revenues	%	decrease
shares	credit	products
approximately	by	ended
settlement	profit	debt
agreement	price	with
operating	rate	activities
quarter	letters	acquired
sale	software	during
increased	contract	an
working	operations	have
expenditures	or	these
at	on	new
as	prices	businesses
company	development	is
customers	estimate	increase
decreased	for	its
fiscal	inventory	costs
s	may	through
net	existing	result
it	gross	current
no	compared	management
valuation	future	changes

(The appendix is continued on the next page.)

Brown, Crowley and Elliott 2020—*JAR*
What Are You Saying? Using topic analysis to Detect Financial Misreporting
False positive rate 49%

TABLE 3

Out-of-Sample Prediction Analysis of topic *and* F-score

Panel A: AUC statistics (AAERs)

Prediction model	AUC
F-score	0.708***
topic	0.680***
topic and *F-score*	0.742***

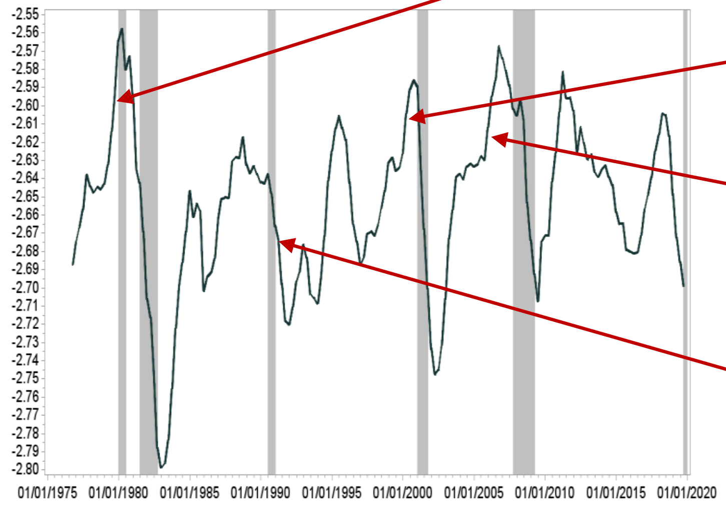
An estimate of the likelihood of misreporting based on M-Score enhances predictions of recessions and economic downturns.

- Premises (e.g., Bernanke and Gertler 1989; Stein 2003; Sadka 2006; Povel, Singh and Winton 2007; Kedia and Philippon 2009; Beatty, Liao and Yu 2013):
  - Misreporting is more likely in periods of expansion because of lax monitoring.
  - Misreporting has real effects because it represents misinformation on which firms base their investment, hiring and production decisions.
- Argument: (Beneish, Farber, Glendenning, Shaw, 2023 TAR)
  - Peers of misreporting firms respond to a decline in economic activity with a delay because they perceive a continuing expansion rather than the true state of the economy.
  - When non-misreporting peer firms ultimately recognize that an economic downturn is occurring, they realize the suboptimal nature of their decisions and curtail their investment and production activity.

Aggregate M-Score and Recessions

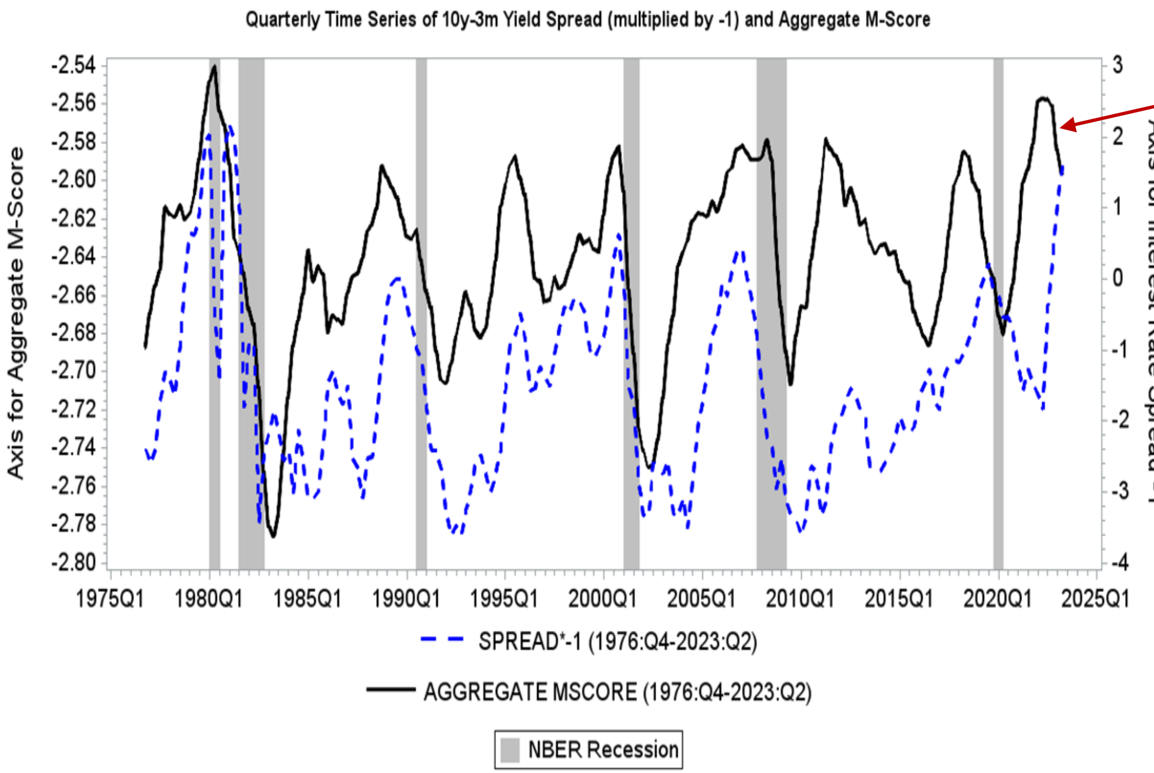
Figure 1
Time Series of Aggregate Misreporting Measures

Panel A: Time Series of $MSCORE_q$



- In the eight quarters preceding the recession that begins in 1980:Q2, aggregate M-Score rises so that the likelihood of aggregate misreporting in the economy increases from 0.41% to 0.52%, **a 26% increase**.
- In the eight quarters preceding the recession that begins in the 2001:Q2, the likelihood of aggregate misreporting rises from 0.43% to 0.48%, **a 12% increase.**
- A similar pattern precedes the 2008:Q1 recession, where the M-Score suggests **a 9% increase** in the likelihood of aggregate misreporting.
- **On the other hand,** for the recession that begins in the 1990:Q4, aggregate M-Score declines slightly from -2.64 to -2.63 which corresponds **to a decrease** in the likelihood of aggregate misreporting from 0.43% to 0.42%

Aggregate M-Score And Yield Spread



- **Probability of a Recession occurring from:**

- **2024:Q1 to 2024:Q4 69.08%**
- **2024:Q2 to 2025:Q1 62.03%**
- **2024:Q3 to 2025:Q2 59.22%**

Questions