The Cost of Fraud Prediction Errors

Messod D. Beneish, Patrick Vorst

The Accounting Review, 2021

叶鑫 2022/01/09

Background

- Costs have long been central to accounting, auditing, and financial economics research as economic agents trade off costs and benefits in making decisions.
- However, empirical documentation of the magnitude of many costs has been scarce. This scarcity is more acute in the context of models predicting financial statement fraud as costs vary not only by decision-maker, but also by type of errors.

		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)
	Non-Fraud	False Negative (False negative cost is incurred whether model is used or not)	True Negative (No cost incurred whether model is used or not)
Precision		True positive /Predicted Fraud	
Sensitivity or recall		True positive / Total Fraud	

Motivation

- First, comparing models by assessing their net costs or benefits is important because thus far, estimates of the rate at which decision makers trade off the costs of false positives and false negatives has remained in the domain of assumptions.

- Second, the most commonly used comparison metrics (ROC-AUC) and expected costs of misclassification (ECM) - either overestimate model performance in studies of rare events and/or make unrealistic assumptions about the relative costs of prediction errors.

- Unlike traditional measures, a cost-based measure can be estimated specifically for each decision maker we consider.

Research Frame

	Auditors	**Investors**	**Regulators**
False Positive Costs (Incorrectly Flagged Non-Fraud Firms)	Incremental audit work **(AT COST)**, lost audit fees resigning from the audit client	Incremental audit work **(AT MARKET)**	If investigations into false positives are made **public**: Costs estimates range from 1 to 15% of the market value of equity.
	Discovery Costs : the estimated probability of litigation	**Loss avoided (profit foregone)** by notinvesting in false positive firms	
False Negatives Costs (Missed Fraud Detection)	**Litigation Costs**	**Loss on the market:** Abnornal Returns in period varying from days -1 to +1 to days-1 to +252 relative to the first public revelation of the fraud times the stock's market value on day -2.	The difference in value-weighted market returns on revelation days over the period 1982-2016 times the value of the stock market in aggregate two days prior to the fraud-revealing announcements.
	Reputation Losses		
	Client Loss		

Fraud Predicton Models

- **The Beneish M-Score (**-1.78**):**
  - $$M\text{-}SCORE = -4.840 + 0.920DSRI_{it} + 0.528GMI_{it} + 0.404AQI_{it} + 0.892SGI_{it} + 0.115DEPI_{it} - 0.172SGAI_{it} + 4.679TATA_{it} - 0.327LGVI_{it} \quad (1)$$
- **The Cecchini (2010) Financial Kernel Model:** support vector machines
- **The Dechow et al. (2011) F-Score (2.45):**
  - $$F\text{-}SCORE = -6.789 + 0.817RSST_{it} + 3.230\Delta REC_{it} + 2.436\Delta INV_{it} + 0.122\Delta CSales_{it} - 0.992\Delta Earnings_{it} + 0.972Issuance_{it} \quad (3)$$
- **The Amiram et al. (2015) FSD Score:** the distribution of first digits
- **The Chakrabarty et al. (2019) Adjusted Benford Score**
- **The misrepresentation model from Alawadhi et al. (2019):**
  - combines variables in previously used models and a set of new variables
- **The machine learning-based prediction from Bao et al. (2020)**

Research Conclusion

- First, for auditors, there are only two economically viable models/decision rules: (i) investigating firms with F-Scores greater than 2.45, (ii) investigating restatement firms with M-Scores greater than 1.78.

- Second, for investors, only the M-Score and, at higher cut-offs, the F-Score provide a net benefit. As these models and/or cutoffs also have the lowest false positive rates, we interpret this result as consistent with the notion that a large number of false positives makes model usage costly.

- Third, we find that the models in Beneish, Dechow et al., Amiram et al. and Chakrabarty et al. are economically viable at low costs for regulators.

- Fourth, we find the usefulness of prediction models within classes of firms that ex-ante have the potential for higher misreporting risk and/or costs. (young growth firms, low operating cash flows, large investing cash outflows, extreme accruals.)

Sample Selection

Our tests are based on three samples of **non-financial** firms:

- Firms that were charged by the SEC with accounting violations other than violations of the Foreign Corrupt Practices Act (hereafter AAER or fraud firms),
- Firms that have restated their financial statements,
- Firms that have a severe restatement, defined as restatements disclosed in press releases or 8-Ks (Huang and Scholz (2012)).

Finally, there are 768 fraud firm-years (313 unique fraud firms), 5408 firm-years (2391 unique restatements) and 2869 firm-year observations corresponding to 1115 unique severe restatements over the period 1979-2018.

Market reaction to the frauds

- Not surprisingly, the mean abnormal return for fraud case is the most adverse (16.1%), followed by the response to the announcement of severe restatements (-2.8%), and to that of all restatements (-1.3%).
- Correspondingly, the market value losses are also more adverse averaging $446.6 million for frauds, and $47.92 ($18.49) million for severe (all) restatements.

Panel B: Three-day Market reaction

Market-Adjusted Abnormal Returns	N	Mean (%)	Mean ($)	Median %	Median $	Sum ($)
Fraud Cases	313	-16.15%	-$446.63	-10.61%	-$20.61	-$139,794.72
All Restatements	2,391	-1.30%	-$18.49	-0.66%	-$0.95	-$44,209.59
Severe Restatements	1,115	-2.79%	-$47.92	-1.50%	-$2.47	-$53,430.80

Traditional Measures

- On average, a firm has fraudulent reports for two consecutive years, focusing on the first year is consistent with the notion that frauds so detected can lead to cost avoidance for all decision makers.

- While investors would benefit from a fraud detection flag any time before the fraud is publicly revealed, a model flagging a firm in its second or in a latter fraud year is unlikely to reduce the costs borne by auditors or regulators.

Panel B: AAERs - Unique Observations

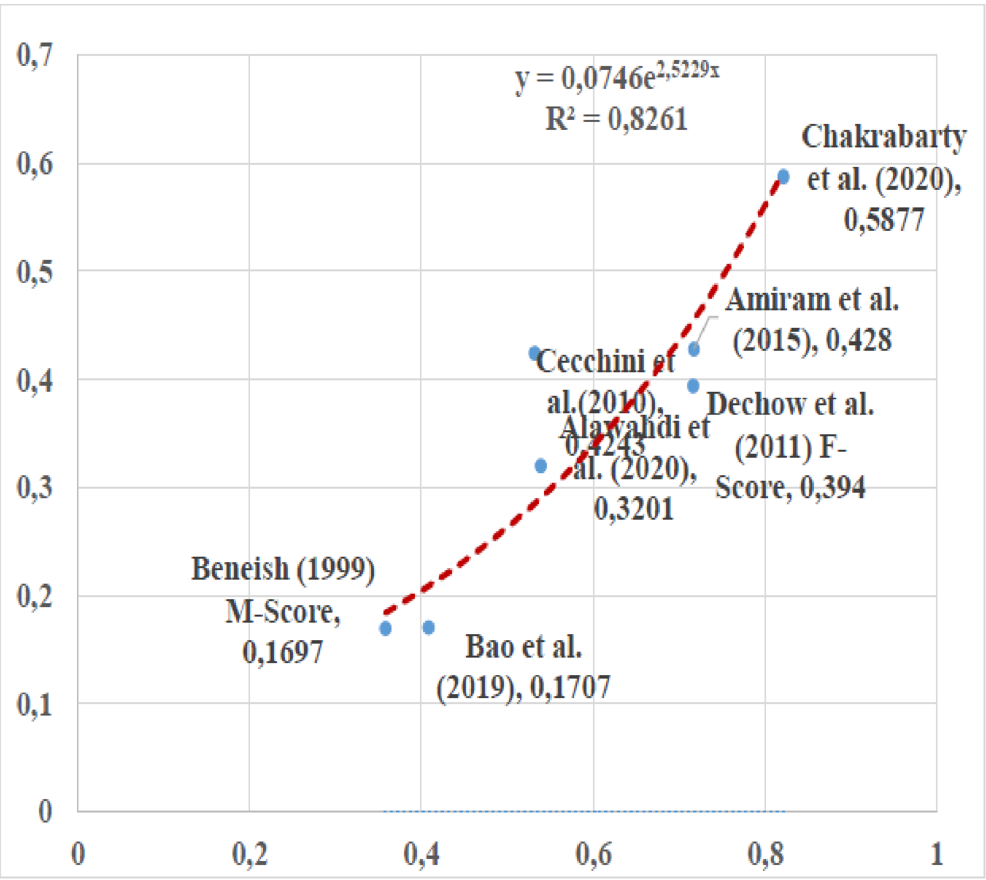
	Fraud Observations	Non-Fraud Observations	AUC	Sensitivity	Precision	False Positive Rate
Beneish (1999) M-Score	313	136,144	0.668	35.78%	0.48%	16.97%
Cecchini et al. (2010)	252	102,273	0.562	53.17%	0.31%	42.43%
Dechow et al. (2011) F-Score	313	136,144	0.717	71.57%	0.42%	39.40%
Amiram et al. (2015) FSD Score	311	131,978	0.709	71.70%	0.39%	42.83%
Alawahdi et al. (2019)	219	106,281	0.691	53.88%	0.35%	32.01%
Bao et al. (2019)	289	115,948	0.537	40.83%	0.59%	17.07%
Chakrabarty et al. (2019) ABF Score	313	134,797	0.713	82.11%	0.32%	58.77%

The relation between false and true positives

- The trend with the best fit is a linear pattern with an R2 of 84% and with a slope coefficient that indicates that a one percent increase in the true positive rate comes at a slightly greater percentage increase in the rate of false positives, an effect magnified by the low proportion of fraud firms in the population.

- Overall, the evidence of a strong positive relation suggests that true positives come at the cost of a greater number of false positives.

Figure 3: False Positive Rates versus True Positive Rates



A Cost-based Measure of Model Performance

- ROC-AUC treats false positives and false negatives as equally costly.

$$FPR = \frac{FP}{FP + TN} \qquad TPR = \frac{TP}{TP + FN}$$

- ROC-AUC focuses on model sensitivity (flagged frauds relative to all frauds), rather than model precision (flagged frauds relative to all firms flagged).

- This leads to potentially misleading conclusions about the relative performance of models because false positives and false negatives are not equally costly, and because AUC measures can mask poor performance when analyzing imbalanced binary classes. (Davis and Goadrich 2006; Ozene et al. 2015)

- For these reasons, we turn to evaluating the costs and benefits of using each model for different decision makers.

The Cost of Classification Errors for Auditors

- **Net benefit (cost)** of using a model: the benefit of TP(avoiding the costs of FN) - the costs of FP

- **Auditors' FN Costs**
  - **litigation component:** actual damages paid by auditors in lawsuits filed against the auditor.
  - **client losses:** the abnormal number of clients lost * the average audit fee per client.
  - **reputational losses:** contagion effect

- **Auditors' FP Costs**

$$E_{AUDit}[FP\ COST] = p\ (AUDFEE_{ijt} * pm\%) + (1\text{-}p)\ (INCRAUDFEE_{ijt} * (1\text{-}pm\%))$$

  - **Discovery cost:** the estimated probability of litigation against an auditor and the estimated settlement relying on models motivated by Honigsberg et al. (2020).

Net Benefits (Costs) to Auditors

- Auditors investigating firms with **F-Scores** greater than **2.45** would gain net benefits across all three samples.
- Investigating **restatement** firms with **M-Scores** greater than **-1.78** also provides net benefits to auditors.

Panel A: AAERs

	True Positives			*False Positives*					*Net Benefits (Costs)*	
Models/Cutoffs	N	Mean Benefit (Avoiding litigation, client, and reputation costs)	Total Benefit (Avoiding litigation, client loss, and reputation costs)	N	Mean Cost (Discovery + Extra Audit Inv.)	Mean Cost (Discovery)	Total Cost (Discovery+ Extra Audit Inv.)	Total Cost (Discovery)	All benefits less all costs	All benefits less discovery costs
Mscore	112	$127.21	$14,247	23,101	$0.88	$0.77	$20,363	$17,842	-$6,116	-$3,595
Fscore > 1	224	$30.95	$6,933	53,643	$1.22	$0.96	$65,362	$51,488	-$58,429	-$44,555
Fscore > 1.85	93	$88.77	$8,256	12,590	$0.98	$0.84	$12,375	$10,532	-$4,119	-$2,276
Fscore > 2.45	56	$311.66	$17,453	5,431	$0.95	$0.83	$5,143	$4,526	**$12,310**	**$12,927**
Cecchini et al.	134	$131.33	$17,598	43,391	$1.44	$1.15	$62,274	$50,064	-$44,676	-$32,466
Amiram et al.	223	$59.31	$13,225	56,532	$1.29	$1.00	$72,907	$56,685	-$59,682	-$43,460
Alawahdi et al.	118	$487.76	$57,556	34,023	$2.61	$2.01	$88,939	$68,373	-$31,383	-$10,817
Bao et al.	118	$33.37	$3,937	19,793	$1.59	$1.23	$31,481	$24,338	-$27,544	-$20,401
Chakrabarty et al.	257	$137.20	$35,260	79,217	$1.42	$1.14	$112,785	$90,158	-$77,525	-$54,898

The Cost of Classification Errors for Investors

- **Investors' FN Costs**
  - **Loss avoided:** avoiding the investment loss associated with the discovery of the misreporting

- **Investors' FP Costs**
  - **the incremental audit fee:** which we assume is billed by the incumbent auditor or required by the newly hired auditor

  - **the profit foregone:** by not investing over the next year in firms flagged by the prediction model as misreporting firms.

Net benefits to investors

- We report results after having winsorized extreme returns to allay concerns that our results are driven by either extreme observations.
- Analyzing the sample of all restatements, we find that investors only benefit when using either the M-Score or the F-Score at higher cut-offs.

Restatements

	True Positives				*False Positives*				*Net: TPB - FPC*	
Models/Cutoffs	N	Mean Benefit (%)	Mean Benefit ($)	Mean Benefit $ (Winsorized)	N	Mean Cost (%)	Mean Cost ($)	Mean Cost $ (Winsorized)	Net Benefit (Cost)	Net Benefit (Cost) Winsorized
Mscore	356	2.19%	$ 20.31	$ 14.41	5,916	3.65%	$ -12.86	$ -4.44	$83,281	$31,410
Fscore > 1	1,060	1.62%	$ 16.92	$ 16.53	18,248	5.27%	$ 56.85	$ 49.73	-$1,019,462	-$889,961
Fscore > 1.85	245	3.70%	$ 19.22	$ 26.27	3,114	-0.53%	$ -39.80	$ -12.63	$128,648	$45,758
Fscore > 2.45	95	3.35%	$ 10.96	$ 32.09	974	-5.22%	$ -228.89	$ -69.92	$223,980	$71,147
Cecchini	1,026	1.06%	$ 43.11	$ 26.75	20,043	8.28%	$ 53.44	$ 38.72	-$1,026,939	-$748,713
Amiram	1,091	1.80%	$ 48.07	$ 20.31	19,356	6.15%	$ 55.14	$ 61.12	-$1,014,767	-$1,160,877
Alawahdi	1,011	1.36%	$ 22.17	$ 20.56	19,957	3.14%	$ 30.11	$ 33.78	-$578,399	-$653,383
Bao	437	1.55%	$ 32.88	$ 24.97	7,728	5.58%	$ 85.97	$ 60.55	-$649,989	-$457,055
Chakrabarty	1,494	1.50%	$ 29.66	$ 20.16	25,638	6.67%	$ 57.39	$ 55.95	-$1,426,919	-$1,404,323

The Cost of Classification Errors for Regulators

- **Investors' FN Costs:** the cost of losing investors' trust in the functioning of capital markets:
  - We estimate this cost as the market-wide dollar abnormal returns in three periods centered on the fraud-revealing announcement (day 0, days -1 to +1, and days -2, +2).

Panel A: False Negative Costs/True Positive Benefits: Announcements and Market-Wide Returns

Return on market-index on Indicator for Revelation date				
Value-weighted	Coeff.	SE	T-stat	p-value
Intercept	0.00054145	0.00011071	4.89	<.0001
Event	-0.00119	0.00050534	-2.36	0.0185
Equal-weighted				
Intercept	0.00084437	0.00009201	9.18	<.0001
Event	-0.00096486	0.00041997	-2.30	0.0216
Return on market-index on indicator for the three days surrounding the revelation date				
Value-weighted				
Intercept	0.00052891	0.00011608	4.56	<.0001
Event	-0.00033379	0.00031751	-1.05	0.2932
Equal-weighted				
Intercept	0.00083902	0.00009647	8.70	<.0001
Event	-0.00030651	0.00026387	-1.16	0.2454

The Cost of Classification Errors for Regulators

- **Investors' FP Costs:** The loss in market value of the firm for which regulators mistakenly investigated misreporting.

  - ➢ We limit the number of flagged firms investigated by the SEC to the top 50 to top 100 ranks for each model in a given year.

  - ➢ Although we understand regulator aversion to publicly making false accusations (we are not aware of any), investors and financial intermediaries carefully monitor regulators actions. For this reason, we consider estimates of market value losses ranging from 1% to 15%.

  - ➢ If regulators make the investigations public, we assume that falsely identified firms experience a market value loss in the range of 3% to 15%, which is the typical market reaction to comment letters and announcements of SEC investigations and charges.

Net benefits to Regulators

- First, three models are too costly to use regardless of the number of investigations or the magnitude of the loss. These are the Cecchini et al. (2010), the Bao et al. (2020), and the Alawahdi et al. (2019) models.
- Second, if the cost of falsely accusing a firm is 1% or less of its market value, four models - Beneish 1999; Dechow et al. 2011; Amiram et al. 2015; Chakrabaty et al. 2020 - are systematically beneficial
- Third, if the cost of falsely accusing a firm is 5% or more, none of the models are beneficial regardless of the number of investigations.

Amiram et al, Model	Number of Frauds Identified	Total Benefit	Number of False Positives	False Positives Cost (1%)	False Positives Cost (3%)	False Positive Cost (5%)	Net Benefit (Cost) at 1%	Net Benefit (Cost) at 3%	Net Benefit (Cost) at 5%
Investigating top 50 per year	20	74,200	1830	34,098	102,294	170,489	40,102	(28,094)	(96,290)
Investigating top 60 per year	25	96,672	2195	40,514	121,542	202,570	56,158	(24,870)	(105,898)
Investigating top 70 per year	30	126,843	2560	45,316	135,948	226,581	81,527	(9,105)	(99,737)
Investigating top 80 per year	31	134,007	2929	51,403	154,210	257,017	82,604	(20,203)	(123,010)
Investigating top 90 per year	32	136,607	3298	60,897	182,690	304,483	75,711	(46,082)	(167,875)
Investigating top 100 per year	37	162,113	3663	67,897	203,691	339,485	94,216	(41,578)	(177,372)

Conclusion

- This paper estimates the costs of fraud prediction errors from the perspective of auditors, investors, and regulators, and proposes a cost-based measure for model comparison.
- We find that the higher true positive rates in recent models come at the cost of higher false positive rates, and that even the best models trade off false to true positives at rates exceeding 100:1.
- For investors, M-Score and the F-Score when used at higher cut-offs are the only models providing a net benefit.
- Overall, as the number of false positives increases, the use of fraud prediction models becomes more costly. Hence, our evidence suggests that researchers should focus on lowering the false positive rates of their models rather than pursuing higher true positive rates.