

Fraud Detection

1:

We applied three models in this study Jones, Modified Jones and Teoh et al Models. This bellowed description table analysis across `jones_eq`, `modified_jones_eq`, and `Teoh_eq` metrics for restating and non-restating firms, including those specifically undergoing a second investigation, uncovers distinct financial reporting patterns. Notably, firms investigated for a second time exhibit markedly lower averages in these metrics, suggesting deeper issues or more pronounced financial adjustments. The reduced variability in these second-investigated firms points to a more uniform set of behaviors or adjustments within this group. Moreover, the presence of significant outliers across all categories highlights extreme cases of financial adjustments, underscoring the complexity and severity of issues leading to restatements and subsequent investigations.

	All Restating jones_eq	All Restating modified_jones_eq	All Restating Teoh_eq	Non-Restating jones_eq	Non-Restating modified_jones_eq	Non-Restating Teoh_eq	Only SEC Restating jones_eq
count	17253.00000	17253.00000	17253.00000	30080.00000	30080.00000	30080.00000	199.00000
mean	2.14297	1.72279	1.78919	1.81861	1.69925	1.59416	1.31402
std	32.66022	33.00495	33.51166	18.19090	35.79965	26.88954	4.43288
min	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00619
25%	0.06515	0.04886	0.04150	0.08581	0.06194	0.05739	0.08500
50%	0.21139	0.16776	0.14218	0.26950	0.20724	0.18465	0.24118
75%	0.73200	0.55151	0.49746	0.82349	0.63638	0.58673	1.20636
max	2926.95089	3245.02730	3245.24752	1627.17871	4409.67468	3622.69026	57.38179

	Only SEC Restating modified_jones_eq	Only SEC Restating Teoh_eq	Non-SEC Restating jones_eq	Non-SEC Restating modified_jones_eq	Non-SEC Restating Teoh_eq
	199.00000	199.00000	17054.00000	17054.00000	17054.00000
	0.78031	0.77111	2.15264	1.73379	1.80107
	1.75082	1.71493	32.84664	33.19627	33.70594
	0.00383	0.00085	0.00000	0.00000	0.00000
	0.05456	0.05764	0.06494	0.04885	0.04135
	0.16002	0.15253	0.21106	0.16792	0.14205
	0.56721	0.49593	0.72657	0.55083	0.49746
	14.33921	11.86630	2926.95089	3245.02730	3245.24752

Moreover, we performed the Mann-Whitney U test results highlight significant disparities in earnings quality metrics (`jones_eq`, `modified_jones_eq`, and `Teoh_eq`) between restating and non-restating firms, with extremely low p-values indicating pronounced differences in financial reporting practices or earnings management. These differences underscore potential concerns about financial reporting quality and manipulation among restating firms. Conversely, when

comparing non-second investigated restating firms to those subjected to second investigations, only `jones_eq` showed a marginal but significant difference, suggesting slight variations in the nature of restatements that attract second scrutiny. However, no significant differences were observed for `modified_jones_eq` and `Teoh_eq`, indicating that these metrics do not capture distinctly different behaviors or adjustments between the two groups of restating firms. This analysis reveals the substantial impact of restatements on earnings quality, while suggesting that second investigations might not always correlate with fundamentally different financial reporting behaviors.

Restating vs. Non-Restating

jones_eq:

U-statistic: 243384648.5, P-value: 2.236456196437614e-29

modified_jones_eq:

U-statistic: 245395850.5, P-value: 7.03451875281237e-23

Teoh_eq:

U-statistic: 240731475.0, P-value: 2.984347906730878e-39

Only Second Investigated Restating vs. Non-Second Investigated Restating

jones_eq:

U-statistic: 1545379.0, P-value: 0.03010524770128475

modified_jones_eq:

U-statistic: 1668111.0, P-value: 0.6805338038646805

Teoh_eq:

U-statistic: 1632433.0, P-value: 0.3562768601324089

2:

There are 8 new features added to each logistic regression models.

1. Prior Stock Performance: *prcc_f* (Price Close - Fiscal) could serve as a proxy for the firm's stock performance. Changes in this variable over time might indicate performance trends.
2. Need for Financing:
 - *dltt* (Long-Term Debt - Total) and *dlc* (Debt in Current Liabilities) can indicate a firm's reliance on external financing.
 - *mkvalt* (Market Value Total) might also give insights into the firm's market valuation and potential financing needs.
3. Company Characteristics (*industry*, *company_age*): These factors provide context on the operational environment and the maturity of the firm. The industry classification helps understand sector-specific risks and practices, while the company's age can indicate its stage in the business lifecycle, each influencing the company's approach to financial reporting and the potential for restatement.

4. Operational Metrics (*sale*): Revenue is a key indicator of operational success but is also a common target for manipulation.
5. Company Identifier (*gvkey*)

Model Summary:

1. Jones:

- Old model:

jones_eq Coefficient is positive but with a p-value of 0.222, indicating a non-significant relationship with the likelihood of restatement at traditional significance levels.

- New model:

- Pseudo R-squared, LLR p-value and Log-Likelihood all improved, suggesting the model is a better fit to the data compared to the old model.
- Significant Predictors: The new model identifies several significant predictors, including *gvkey, industry_freq, dltd, dlc, mkval, and company_age*, each with a p-value well below 0.05, indicating strong evidence of their association with the likelihood of restatement. Notably, *jones_eq* becomes significant in this model (p-value: 0.021).
- Accuracy: The new model trades off a slight decrease in accuracy and precision for significant gains in recall and F1 score, indicating a much better balance in identifying true restatements. Despite slightly lower precision, its enhanced recall makes it more practical and effective for predicting financial restatements compared to the old model.

2. Modified Jones:

- Old model:

The *modified_jones_eq* Coefficient, showing a p-value of 0.986, indicates a statistically non-significant relationship with the likelihood of restatement. This suggests that, in isolation, the modified Jones model does not effectively predict restatements.

- New model:

- The model shows enhancements in model fit indicators, such as Pseudo R-squared, LLR p-value, and Log-Likelihood, implying a more accurate representation of the data than the simpler old model.
- Significant Predictors: Improved model complexity reveals key significant predictors including *gvkey, industry_freq, dltd, dlc, mkval, and company_age*, all demonstrating a statistically significant relationship with restatement likelihood (p-values below 0.05). Interestingly, within this expanded context, modified_jones_eq still does not reach traditional levels of statistical

significance (p-value: 0.147), suggesting its limited predictive power on restatements when compared to other firm-level factors.

- Accuracy: The old model, despite a seemingly decent accuracy, is practically ineffective due to its failure to identify any actual restatement cases. The new model demonstrates a more balanced and effective approach, with substantial improvements in recall and F1 score, making it significantly more useful for practical applications, despite a minor decrease in accuracy. The ability to identify true restatements, as shown by the recall and F1 score improvements, marks a significant advancement in predictive capability.

3. Teoh et al Model:

- Old model: The *Teoh_eq* coefficient is slightly positive but not statistically significant (p-value: 0.549), indicating that the Teoh model, on its own, does not have a significant predictive relationship with the likelihood of financial restatement.
- New model:
 - An improvement in overall fit and predictive power, as evidenced by better values for Pseudo R-squared, LLR p-value, and Log-Likelihood compared to the old model. This suggests the comprehensive model is better suited to capturing the complexities around financial restatements.
 - Significant Predictors: key predictors emerge as significant, including *gvkey, industry_freq, dltd, dlc, mkval, and company_age*, all showing a strong statistical relationship with the restatement likelihood (p-values below 0.05). *Teoh_eq*, while showing a positive coefficient, remains not statistically significant (p-value: 0.132), suggesting its limited direct influence on restatement likelihood in the presence of other firm-level factors.
 - Accuracy: Transitioning from the old to the new model demonstrates a negligible accuracy dip for significant recall and F1 score improvements. Despite lower precision, the new model's enhanced recall markedly boosts its practicality, making it far more effective in identifying financial restatements and thus, a preferable choice for predictive analysis.

3: The 100 firms that are most likely to restate earnings in the future are stored in the *top_100_firms*, *top_100_firms_modified_jones*, and *top_100_firms_teoh*.

1. Jones model:

The model accuracy is 61%. In general, out of 100 firms, there are 52 firms identified as restatement in 2019, 53 firms identified as restatement in 2020 and 51 firms identified as restatement in 2021.

Among the top 100 firms regarding financial statement restatements over the three years studied. A significant portion, 52 firms, did not restate their financial statements at all, indicating a level of financial reporting stability or accuracy within this group. On the other end, 41 firms restated their financial statements all three years, suggesting persistent issues with financial accuracy or reporting practices in these entities. A small minority, 7 firms in total, show intermediate behavior with one or two restatements, pointing to occasional discrepancies in financial reporting. This bifurcation suggests distinct groups within the top 100 firms: one demonstrating consistent reporting accuracy and another facing ongoing challenges in financial statement reliability.

2. Modified Jones model:

The model accuracy is still 61%. In general, out of 100 firms, there are 67 firms identified as restatement in 2019, 63 firms identified as restatement in 2020 and 61 firms identified as restatement in 2021.

It reveals a near-even split between firms with no restatements and those restating all three years, 40 and 54 firms respectively. This polarization suggests the model effectively identifies firms with consistent financial reporting practices as well as those with recurring restatement issues. The minimal number of firms (6 in total) with one or two restatements indicates fewer instances of intermittent reporting discrepancies, suggesting that firms tend to either consistently meet reporting standards or persistently face challenges, rather than fluctuating between these states. This analysis underscores the importance of targeted interventions for the latter group to improve financial reporting accuracy.

3. Teoh et al model:

The model accuracy is still 61%. In general, out of 100 firms, there are 67 firms identified as restatement in 2019, 63 firms identified as restatement in 2020 and 61 firms identified as restatement in 2021.

Result is the same as Modified Jones model, a near-even split between firms with no restatements and those restating all three years, 40 and 54 firms respectively.

4:

New Model's performance on a training dataset (from year 2010 to year 2020) of 50,923 observations shows a modest accuracy of 62.01%, with significant limitations highlighted by a recall of only 10.70%. This indicates the model's restricted capacity to accurately identify firms likely to restate their financial statements. Despite a reasonable precision of 54.68%, the low recall and F1 score of 17.90% suggest a substantial number of actual restatements are missed, pointing towards an imbalance in the model's predictive ability. Significant predictors such as 'industry_freq', 'dlc', 'company_age' and 'total_EM_index' affirm their influence on restatement likelihood (since their p-value are less than 0.05). However, the model's overall modest explanatory power and predictive performance underline the necessity for further refinement and exploration of additional variables to enhance its predictive accuracy in identifying financial restatements.

```
Optimization terminated successfully.
Current function value: 0.655363
Iterations 5
```

Logit Regression Results

```
=====
Dep. Variable:      restatement    No. Observations:      50923
Model:              Logit         Df Residuals:          50913
Method:             MLE          Df Model:              9
Date:               Sun, 07 Apr 2024  Pseudo R-squ.:         0.01801
Time:               20:04:33         Log-Likelihood:        -33373.
converged:          True           LL-Null:              -33985.
Covariance Type:    nonrobust      LLR p-value:          6.299e-258
=====
```

	coef	std err	z	P> z	[0.025	0.975]
const	-0.8077	0.033	-24.192	0.000	-0.873	-0.742
gvkey	-1.067e-06	1.31e-07	-8.123	0.000	-1.32e-06	-8.09e-07
sale	1.06e-06	9.53e-07	1.113	0.266	-8.07e-07	2.93e-06
industry_freq	-2.2273	0.154	-14.454	0.000	-2.529	-1.925
prcc_f	9.75e-06	5.63e-05	0.173	0.862	-0.000	0.000
dltt	2.482e-06	3.88e-06	0.640	0.522	-5.12e-06	1.01e-05
dlc	-4.263e-05	1.46e-05	-2.914	0.004	-7.13e-05	-1.4e-05
mkvalt	-4.529e-07	4.51e-07	-1.005	0.315	-1.34e-06	4.31e-07
company_age	0.0322	0.001	23.291	0.000	0.030	0.035
total_EM_index	0.0107	0.001	8.695	0.000	0.008	0.013

```
=====
```

The model's performance on testing dataset (year 2021), with an accuracy of 62.01%, precision of 54.68%, recall of 10.70%, and an F1 score of 17.90%, reveals significant challenges in its predictive efficacy. The model is notably conservative, indicated by its low recall, missing a large proportion of actual restatements (Type 2 errors) while also generating a considerable number of false alarms (Type 1 errors) as shown by its precision rate. This imbalance suggests a crucial need for model refinement to enhance its ability to accurately predict financial restatements. Improving the model could involve better feature selection, exploring more sophisticated algorithms, or adjusting the classification threshold to achieve a more balanced trade-off between sensitivity and specificity, thereby reducing both types of errors and increasing the overall predictive value.

Here is a list of 10 firms that most likely to restate financials, predicted by the model:

gvkey	tic
7435	MMM
3246	CMC
2111	BDX
2080	BSET
1004	AIR
1327	SWKS
2710	STZ
7146	MKC
8850	KWR
1632	ADI

Now let's narrow down the list to 5 firms you believe have the highest potential to restate 2021 financials:

1. 3M (MMM):

3M operates across diverse sectors with complex supply chains and regulatory requirements. Given its involvement in various legal and environmental challenges over the years, the complexity of its operations could potentially lead to accounting errors or necessitate adjustments upon further review.

Before 2021, one of the significant issues 3M faced was related to its Bair Hugger surgical warming blankets. This device came under legal scrutiny due to allegations that it could cause infections post-surgery. The allegations centered around the device's design, which was said to allow for the introduction of contaminants into the surgical site. The concern was that warmed air, rising against the downward airflow in the operating room, could deposit bacteria from non-sterile portions of the operating room onto the surgical site. These lawsuits progressed to a Multidistrict Litigation (MDL) stage, with more than 800 cases consolidated under a judge in Minnesota. The litigation process included selecting bellwether cases, which are representative cases used to test the waters for the upcoming trials. This legal challenge was significant as it pointed to potential risks associated with one of 3M's medical products, emphasizing the scrutiny the company was under regarding the safety and design of its products (<https://www.natlawreview.com/article/bair-hugger-lawsuit-multidistrict-litigation-judge-issues-bellwether-trial-order>).

These legal and regulatory challenges before 2021 highlighted the complex environment in which 3M operated, encompassing product liability and environmental issues. Although these concerns primarily revolved around legal liabilities rather than accounting or financial fraud, they indicated areas of risk that could potentially impact the company's financial statements or lead to financial restatements if these liabilities were not properly accounted for or if they had a substantial impact on the company's financial health.

2. Skyworks Solutions (SWKS): As a semiconductor company, Skyworks is part of a highly dynamic and competitive industry. The rapid pace of technological change and complex revenue recognition issues, especially related to licensing and intellectual property, could lead to a higher likelihood of financial restatement.

The semiconductor industry faced significant supply chain challenges before 2021, partly due to the COVID-19 pandemic. Skyworks, like its peers, might encounter complexities in inventory valuation due to these disruptions. Changes in estimated production costs, obsolescence, or write-downs of inventory could necessitate financial adjustments. Also, the aggressive growth targets and expansion into new technological areas, including 5G and automotive semiconductors, involve making substantial financial projections. If these projections do not materialize as expected, it could lead to goodwill impairments, a common trigger for financial restatements.

3. Constellation Brands (STZ): Involved in alcohol and cannabis industries, Constellation Brands engages in significant M&A activities. The valuation and integration of these deals are complex and subject to significant judgment, which could lead to restatements, especially if market conditions change or assumptions prove inaccurate.

Impairment Charge Related to Mexicali Facility: After the Mexican government's intervention, Constellation Brands had to halt its production facility in Mexicali, leading to a significant impairment charge. This incident highlights the operational and regulatory risks the company faced in its international operations (<https://www.brewbound.com/news/constellation-brands-beer-sales-top-6-billion-in-fiscal-year-2021>).

4. McCormick & Company (MKC): Though perhaps less likely than technology or manufacturing firms to need to restate financials due to the nature of its business, McCormick's global operations and any significant acquisitions could pose challenges in financial reporting, potentially leading to restatements.
5. Analog Devices (ADI): Analog Devices operates in the semiconductor industry, which is characterized by rapid innovation and significant R&D expenditure. The valuation of intangible assets, including patents and technology licenses, as well as revenue recognition practices, might lead to restatements.