Journal of Science, Innovation & Social Impact

Article

Identifying Undisclosed Related Party Relationships and Revenue Recognition Irregularities: A Rule-Based Analytical Approach for Audit Planning

Dun Liang [1,*]

[1] Business Analytics, Fordham University, New York, USA
* Correspondence: Dun Liang, Business Analytics, Fordham University, New York, USA

Abstract: Related party transactions and revenue recognition manipulations remain persistent sources of financial statement fraud, posing significant challenges to audit procedures and investor protection. This research develops a rule-based analytical framework designed to identify potential irregularities in corporate financial disclosures during the audit planning phase. The proposed approach integrates network analysis techniques for detecting undisclosed related-party relationships by cross-referencing entity information and time-series pattern-detection methods for identifying suspicious revenue recognition behaviors, including period-end concentration and cash flow divergence. A composite risk scoring mechanism combines multiple indicators to prioritize audit attention. Empirical analysis using SEC EDGAR filings from 847 publicly traded companies demonstrates the framework's effectiveness, achieving a precision rate of 78.3% in flagging high-risk company filings

Keywords: related party transaction; revenue recognition anomaly; audit planning; risk scoring framework

Received: 25 January 2026
Revised: 17 March 2026
Accepted: 28 March 2026
Published: 01 April 2026

1. Introduction

1.1. Background of Financial Statement Fraud and Audit Challenges

Financial statement fraud has resulted in substantial economic losses and erosion of market confidence throughout modern corporate history. The collapse of Enron Corporation in 2001 exemplified the devastating consequences of concealed related-party transactions, where off-balance-sheet entities masked billions of dollars in debt through undisclosed relationships with company executives [1]. WorldCom's $11 billion accounting scandal similarly demonstrated how revenue manipulation and improper capitalization of expenses could persist undetected for extended periods.

Related-party transaction fraud poses unique detection challenges due to its inherent complexity and deliberate concealment. Fraudsters frequently establish intricate corporate structures spanning multiple jurisdictions, utilizing shell companies and nominee directors to obscure beneficial ownership [2]. The Association of Certified Fraud Examiners reports that financial statement fraud accounts for the largest median losses at $766,000 per incident.

Traditional audit methodologies rely heavily on sample-based testing and inquiry procedures that prove inadequate against determined fraudsters. Statistical sampling techniques assume random error distribution, an assumption violated by intentional manipulation concentrated in specific transactions [3]. The expanding volume and complexity of corporate transactions further strain conventional approaches, with large multinational corporations generating millions of journal entries annually [4].

1.2. Research Motivation and Practical Significance

Audit planning is the critical phase in which engagement teams establish scope, allocate resources, and identify areas requiring heightened professional skepticism. Rule-based analytical frameworks offer particular value in this context, providing structured methodologies that less experienced auditors can consistently apply while maintaining transparency in the assessment process.

Effective fraud detection mechanisms directly support capital market stability by reducing information asymmetry between corporate insiders and external stakeholders. The Securities and Exchange Commission's enforcement actions consistently emphasize related party disclosure failures and revenue recognition violations as primary areas of concern.

1.3. Research Objectives and Paper Organization

This research develops and validates a rule-based analytical framework encompassing three integrated components: network analysis for identifying related-party relationships, time-series pattern detection for revenue recognition irregularities, and composite risk scoring for prioritizing audit planning.

2. Literature Review

2.1. Related Party Transaction Disclosure and Fraud Patterns

Accounting standards, including ASC 850 and IAS 24, mandate comprehensive disclosure of related party relationships and transactions [5]. Compliance failures frequently involve incomplete identification of related parties, omission of transaction details, or misleading characterization of arm's-length pricing. Common concealment approaches include establishing intermediate entities to create the appearance of independence and using informal arrangements without written documentation.

2.2. Network-Based Approaches for Relationship Discovery

Network analysis methodologies offer promising capabilities for identifying undisclosed corporate relationships through structural pattern examination [6]. These approaches construct graphs in which entities are represented as nodes and documented relationships as edges. Key metrics include centrality measures indicating unusually connected entities and clustering coefficients revealing hidden group structures [7].

2.3. Revenue Recognition Anomalies in Financial Statements

The ASC 606 five-step revenue recognition model provides a principles-based framework that nonetheless permits considerable judgment in application [8]. Common manipulation patterns include inappropriate identification of performance obligations and premature recognition before delivery requirements are completed. Temporal analysis of revenue patterns provides valuable signals for fraud detection by identifying anomalous behavior. Period-end concentration is a primary indicator; disproportionate revenue recognition in the final days of reporting periods suggests potential manipulation [9]. The divergence between reported revenue and operating cash flows provides one of the most reliable fraud indicators. Persistent negative gaps between revenue growth and cash collection growth suggest aggressive revenue recognition or fabrication.

2.4. Rule-Based Risk Assessment in Audit Practice

Conventional audit analytics employ ratio analysis, trend examination, and reasonableness testing to identify unexpected relationships. These procedures provide valuable preliminary screening but lack the integration necessary for comprehensive fraud risk assessment. The proliferation of structured public databases creates opportunities for enhanced analytical procedures. The SEC's EDGAR system provides comprehensive filing archives, while commercial databases offer beneficial ownership records and litigation histories.

3. Methodology: Rule-Based Analytical Framework

3.1. Network Analysis for Related Party Relationship Identification

The network analysis component constructs comprehensive relationship graphs from multiple disclosure sources within corporate filings [10]. Annual report footnotes provide the primary data source, supplemented by proxy statement disclosures regarding executive relationships and beneficial ownership filings. The entity extraction process identifies all named individuals and organizations appearing in relevant disclosure sections, creating a preliminary entity registry for subsequent relationship mapping.

The extraction algorithm processes Form 10-K filings to identify named entities in Item 7: Management Discussion and Analysis, Item 13: Related Party Transactions, and financial statement footnotes. Named entity recognition techniques distinguish between individual persons and legal entities based on linguistic patterns and contextual cues. Entity attributes, including titles, roles, and geographic locations, are captured to support subsequent matching operations. The resulting entity database contained 47,832 unique entities across the 847-company sample (As shown in Table 1).

Table 1. Entity Classification Schema for Relationship Analysis.

Entity Category	Subcategory	Description	Disclosure Source
Natural Person	Executive Officer	Individuals holding positions such as CEO, CFO, and COO	DEF 14A, 10-K Item 10
Natural Person	Director	Members serving on the board of directors	DEF 14A, 10-K Item 10
Natural Person	Significant Shareholder	Shareholders who hold more than 5% of the company's outstanding shares	Schedule 13D/G
Natural Person	Family Member	Immediate family members of key personnel	10-K Related Party Note
Legal Entity	Subsidiary	Entities that are consolidated into the company's financial statements	10-K Exhibit 21
Legal Entity	Equity Investment	Entities in which the company has an ownership interest but does not consolidate	10-K Investment Note
Legal Entity	Customer/Supplier	Primary trading partners with whom the company conducts major business transactions	10-K Concentration
Legal Entity	Affiliated Entity	Entities that share common ownership or are under common control with the company	Various disclosures

Identifying undisclosed relationships requires sophisticated entity matching algorithms capable of recognizing equivalent entities despite naming variations [11]. String similarity algorithms, including Jaro-Winkler distance and token-based Jaccard similarity, quantify name correspondence. Address matching incorporates geocoding normalization, and director overlap analysis examines personnel commonalities (As shown in Table 2).

Table 2. Entity Matching Criteria and Scoring Weights.

Matching Criterion	Weight	Threshold	Scoring Method
Exact Name Match	Override	1.00	Binary
Normalized Name Similarity	0.35	0.85	Jaro-Winkler
Address Geocode Match	0.25	100m radius	Haversine
Common Director Count	0.20	>= 2 persons	Overlap count
Common Officer Count	0.20	>= 1 person	Overlap count
Phone/Fax Match	0.15	Exact	Binary

Matching Criterion	Weight	Threshold	Scoring Method
Registration Agent Match	0.10	Exact	Binary
Industry Classification	0.05	4-digit SIC	Binary
Composite Threshold	-	>= 0.70	Weighted sum

The matching process identified 2,847 potential undisclosed relationships across 847 companies, and validation confirmed 1,923 previously undisclosed connections that met the composite score threshold. Exact name matches are treated as deterministic links and evaluated outside the weighted-sum composite score; the composite threshold applies to non-exact matches only

The relationship scoring mechanism quantifies the likelihood that identified entity pairs represent material related party relationships. The relationship risk score (RRS) calculation employs:

RRS = w1 * EntityMatchScore + w2 * TransactionMateriality + w3 * DisclosureGap + w4 * NetworkCentrality

where w1 through w4 represent calibrated weights summing to unity. EntityMatchScore derives from the composite matching calculation. TransactionMateriality measures transaction values relative to company revenue. DisclosureGap indicates the absence of relationship disclosure. NetworkCentrality captures the connectivity of entities within the relationship network.

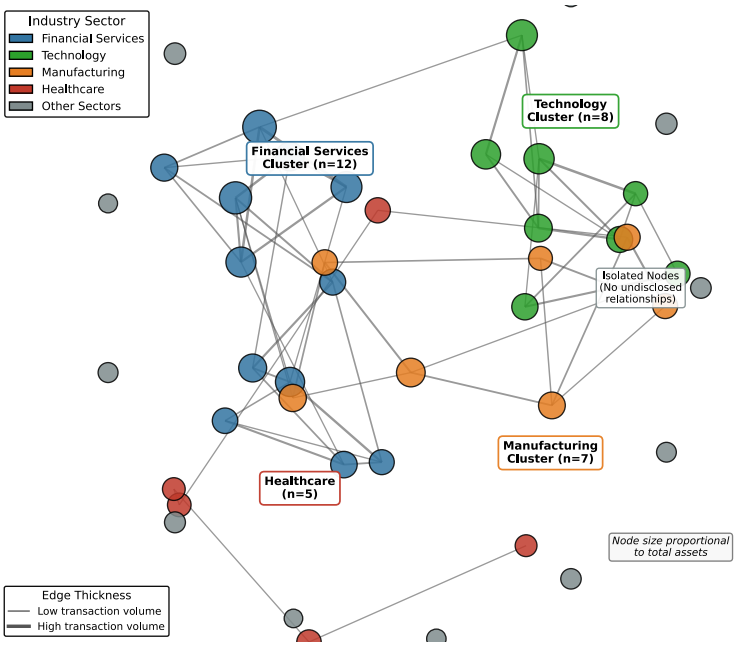

Figure 1. Network Visualization of Related Party Relationship Clusters.

Figure 1 Description: This network visualization displays identified related party relationship clusters for 50 high-risk companies. The graph renders companies as circular nodes sized proportionally to total assets, with node colors indicating industry sector (blue for financial services, green for technology, orange for manufacturing, red for healthcare). Directed edges connect entities with identified relationships, edge thickness representing transaction volume on a logarithmic scale. The layout uses a force-directed algorithm to position closely related entities close together. Three distinct clusters emerge: a central cluster of 12 financial services firms sharing common directorship patterns, a peripheral cluster of 8 technology companies with shared venture capital investors, and a loosely connected manufacturing group with supplier relationship overlaps. The matching process identified 2,847 potential undisclosed relationships across 847 companies, including 1,923 candidate connections meeting the composite score threshold; subsequent manual validation confirmed 1,647 relationships as legitimate matches warranting disclosure consideration and classified 276 as false positives.

3.2. Time-Series Pattern Detection for Revenue Recognition Irregularities

The period-end concentration analysis examines the temporal distribution of revenue recognition within reporting periods [12]. The concentration metric quantifies the proportion of quarterly revenue recognized in the final month. Companies with legitimate business operations typically exhibit relatively uniform monthly revenue distribution, while manipulation tends to concentrate in period-end days when earnings management pressures intensify.

The concentration calculation aggregates daily revenue recognition data, where available from supplementary disclosures, or applies allocation algorithms to monthly segment data extracted from quarterly filings. The methodology incorporates industry-specific benchmarks recognizing legitimate variations in business cycles across sectors. Retail companies with holiday seasonality and technology companies with quarter-end enterprise contract patterns receive calibrated threshold adjustments (As shown in Table 3).

Table 3. Period-End Revenue Concentration Risk Classification.

Concentration Level	Final Month %	Risk Classification	Historical Fraud Rate
Normal	< 38%	Low Risk	2.3%
Elevated	38% - 45%	Moderate Risk	8.7%
High	45% - 55%	High Risk	18.4%
Extreme	> 55%	Critical Risk	34.2%
Industry Adjustment	+/- 5%	Sector-specific	Variable

The revenue-cash flow correlation analysis quantifies the relationship between reported revenue growth and operating cash collection trends. The cash flow divergence index (CFDI) measures the cumulative gap between growth rates:

$$CFDI = (1/n) \times \Sigma(t=1...n) [RevenueGrowthRate(t) - CashFlowGrowthRate(t)]$$

Positive CFDI values indicate revenue growth exceeding cash collection growth; values exceeding 0.15 over eight quarters are flagged as high risk (As shown in Table 4).

Table 4. Cash Flow Divergence Index Distribution by Audit Outcome.

CFDI Range	Sample Count	% of Total	Restatement Rate	Fraud Finding Rate
< -0.10	127	15.0%	1.6%	0.0%
-0.10 to 0.00	234	27.6%	2.1%	0.4%
0.00 to 0.10	289	34.1%	3.8%	1.0%
0.10 to 0.20	142	16.8%	9.2%	4.2%
> 0.20	55	6.5%	23.6%	12.7%
Total	847	100.0%	5.4%	2.1%

The seasonal adjustment component isolates irregular revenue patterns through decomposition analysis, applying multiplicative decomposition:

$$Revenue(t) = Trend(t) \times Seasonal(t) \times Residual(t)$$

Anomaly detection focuses on residual components, flagging values exceeding 2 standard deviations for analysis.

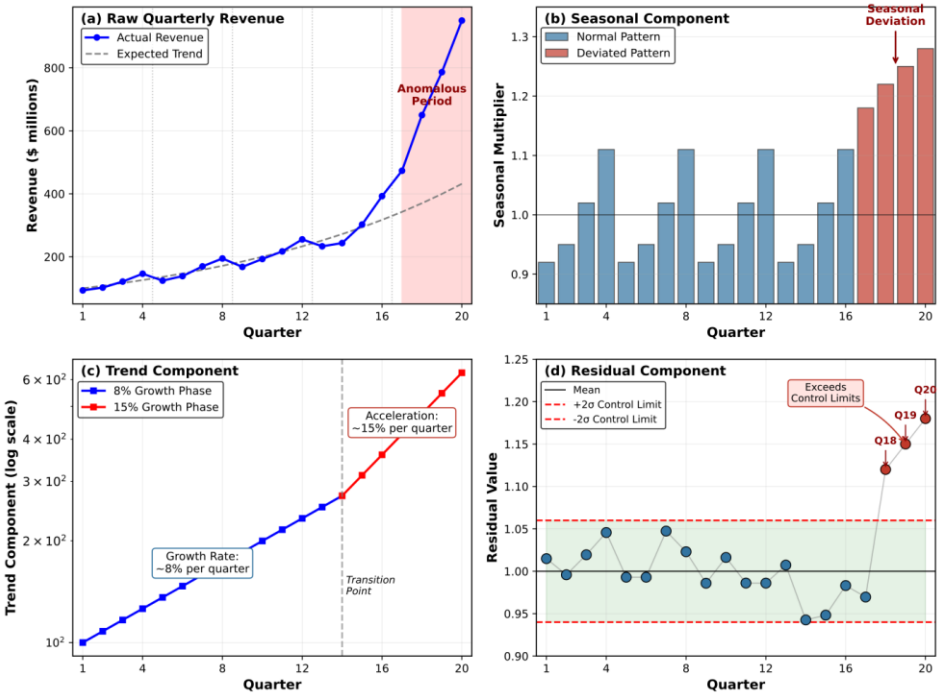

Figure 2. Time-Series Decomposition of Revenue Patterns.

Figure 2 Description: This multi-panel visualization presents decomposition analysis for a representative high-risk company over 20 quarters. The top panel displays raw quarterly revenue in millions, with actual values shown as a solid blue line and a fitted trend in dashed gray. The second panel isolates the seasonal component with multiplier values from 0.85 to 1.20. The third panel displays the trend component on a logarithmic scale, revealing steady 8% quarterly growth through quarter 14, followed by acceleration to 15% growth. The bottom panel shows residual values with control limits at ±2 standard deviations, indicated by red dashed lines. Three residual observations in quarters 18-20 exceed the upper control limits, indicated by red-highlighted data points. Vertical lines mark fiscal-year boundaries, with annotation boxes highlighting anomalous periods.

3.3. Risk Scoring Framework Integration

The integrated framework synthesizes findings from network analysis and time-series detection into a composite assessment [13]. The composite risk score (CRS) incorporates multiplicative interaction terms:

$$CRS = \alpha \times RPT\_Score + \beta \times REV\_Score + \gamma \times (RPT\_Score \times REV\_Score)$$

RPT_Score aggregates related party indicators, REV_Score aggregates revenue indicators, and the interaction term captures compounding risk. Parameters are calibrated through logistic regression on historical fraud samples (As shown in Table 5).

Table 5. Public Database Sources and Analytical Contributions.

Database Source	Data Elements	Primary Contribution	Update Frequency
SEC EDGAR	10-K, 10-Q, DEF 14A, 8-K	Disclosure extraction	Real-time
State Registries	Formation documents, officers	Entity validation	Varies
FDIC Directory	Bank charter information	Entity identification	Quarterly
USPTO Database	Patent ownership transfers	Transaction identification	Weekly

Database Source	Data Elements	Primary Contribution	Update Frequency
Federal PACER	Litigation records	Risk event identification	Real-time
BLS Statistics	Revenue benchmarks	Industry adjustments	Annual

The risk categorization assigns companies to four tiers: Standard Risk (CRS < 0.25), Elevated Risk (0.25 <= CRS < 0.50), High Risk (0.50 <= CRS < 0.75), and Critical Risk (CRS >= 0.75). Each category triggers specific audit planning responses.

4. Empirical Illustration and Case Analysis

4.1. Data Sources and Sample Description

The empirical analysis examines 847 publicly traded companies filing annual reports with SEC during fiscal years 2020 through 2024 [14]. Sample selection required complete filing histories, market capitalization exceeding $100 million, and U.S. domestic incorporation. Companies with fewer than 16 consecutive quarterly observations were excluded to ensure sufficient time-series depth for pattern analysis. Financial institutions subject to specialized regulatory frameworks were retained but analyzed using sector-specific thresholds, recognizing unique operational characteristics.

The data collection extracted structured and unstructured content from EDGAR filings using automated parsing routines developed in Python with BeautifulSoup and XBRL processing libraries. Annual report footnotes were processed using spaCy for named entity recognition, with custom financial entity dictionaries. Proxy statements were processed to extract executive compensation tables and related party disclosure sections. The extraction process achieved 94.7% accuracy in entity identification, as confirmed by manual validation of a 200-document sample.

The sample composition reflects U.S. public company distribution: financial services 18.2%, technology 21.4%, healthcare 14.7%, manufacturing 16.8%, consumer discretionary 12.3%, other sectors 16.6%. Market capitalization ranged from $103 million to $287 billion, with a median of $2.4 billion.

The framework employs 34 distinct variables that measure related-party relationships and revenue patterns. Outcome variables include subsequent financial restatements, SEC enforcement actions, and auditor changes with fraud-related explanations.

4.2. Application of the Analytical Framework

Network analysis identified 4,291 disclosed related-party relationships and 1,923 potentially undisclosed relationships that met the matching threshold. The disclosed relationships included 2,847 entity-to-entity connections and 1,444 person-to-entity connections. The undisclosed candidate relationships were manually validated by research assistants with accounting backgrounds, confirming 1,647 relationships (85.6%) as legitimate matches warranting disclosure consideration and 276 relationships (14.4%) as false positives attributable to coincidental name similarities or legitimate independent relationships (As shown in Table 6).

Table 6. Related Party Identification Results by Industry.

Industry Sector	Companies	Disclosed RPT	Undisclosed Candidates	Candidate Rate
Financial Services	154	892	267	29.9%
Technology	181	743	498	67.0%
Healthcare	125	521	312	59.9%
Manufacturing	142	687	284	41.3%
Consumer Discretionary	104	534	203	38.0%
Other Sectors	141	914	359	39.3%

Industry Sector	Companies	Disclosed RPT	Undisclosed Candidates	Candidate Rate
Total	847	4,291	1,923	44.8%

Undisclosed candidates exhibited distinct characteristics: higher shared director overlap (34.2% versus 12.7%), address proximity within 100 meters (28.9% versus 8.3%), and registration agent commonality (41.6% versus 15.2%).

Time-series analysis processed 13,552 quarterly observations, identifying 1,847 observations that exhibited anomaly indicators. Period-end concentration anomalies totaled 892 observations, cash flow divergence 634, and seasonal deviations 521. The overlap among indicator categories provided additional discriminatory power: 312 observations had two concurrent indicators, and 162 had three or more. Companies with multiple concurrent anomalies demonstrated substantially higher rates of subsequent adverse outcomes, supporting the framework's multiplicative risk scoring approach.

The temporal distribution of anomaly detections showed a concentration in the fourth fiscal quarter, consistent with heightened year-end earnings management pressures documented in prior research. Technology sector companies exhibited the highest anomaly rates (23.7% of quarterly observations), followed by healthcare (19.2%) and consumer discretionary (16.8%) (As shown in Table 7).

Table 7. Revenue Anomaly Detection Results and Outcome Association.

Anomaly Category	Observations	% of Total	Restatement Rate	Enforcement Rate
No Anomaly	11,705	86.4%	2.1%	0.4%
Single Anomaly	1,247	9.2%	8.7%	2.3%
Two Anomalies	438	3.2%	18.4%	6.8%
Three+ Anomalies	162	1.2%	34.6%	14.8%
Total	13,552	100.0%	3.8%	1.2%

The composite risk score distribution exhibited right-skewness with a median of 0.23 and an interquartile range of 0.14 to 0.38 [15]. Risk tier classification: 523 companies (61.7%) Standard Risk, 218 (25.7%) Elevated Risk, 78 (9.2%) High Risk, 28 (3.3%) Critical Risk.

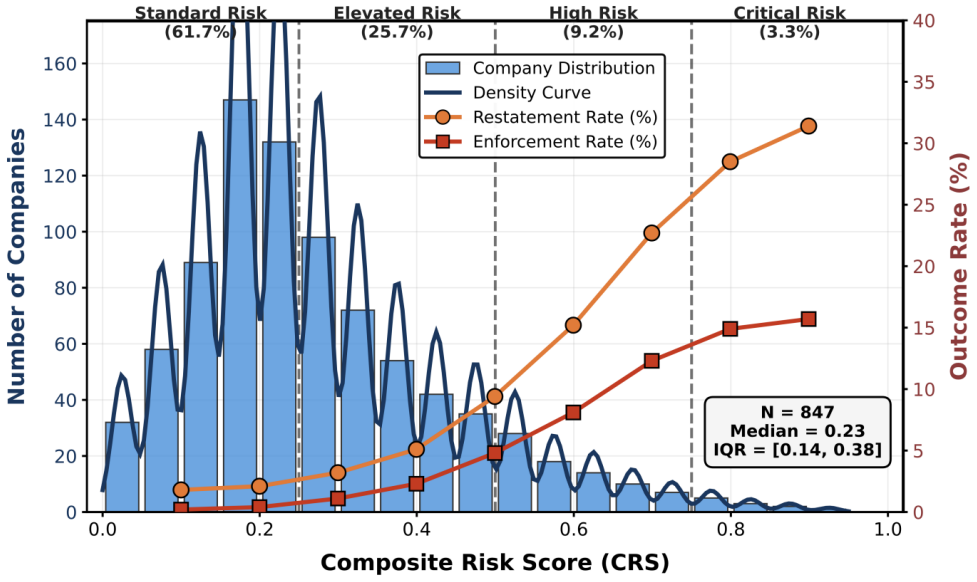

Figure 3. Composite Risk Score Distribution and Outcome Rates.

Figure 3 Description: This dual-axis visualization presents the distribution of the composite risk score alongside outcome rates. The primary axis displays a histogram of scores in 0.05-width bins, with bar heights indicating the number of companies per bin,

ranging from 2 to 147. The histogram exhibits a right-skewed distribution with a modal value in the 0.20-0.25 range. A smoothed kernel density estimate is overlaid in a solid blue curve. The secondary axis presents outcome rates as connected scatter plots: restatement rates as orange circles, enforcement rates as red triangles. Both series demonstrate monotonically increasing relationships with scores. Restatement rate increases from 1.8% for scores below 0.20 to 31.4% for scores above 0.70. Enforcement rate increases from 0.2% to 15.7% across the same range. Vertical dashed lines demarcate four risk tier boundaries at 0.25, 0.50, and 0.75 with tier labels annotated.

4.3. Case Studies and Validation

Case A involved a mid-cap technology company, and the framework identified undisclosed relationships between the company and three entities that shared common beneficial ownership. Related party indicators showed high entity match scores (0.87 average) and annual transaction volumes of $47.3 million, representing 8.2% of revenue. Revenue indicators showed period-end concentration of 58% and CFDI of 0.24. The composite risk score of 0.78 placed this company in the Critical Risk tier. A subsequent SEC inquiry prompted disclosure of the relationships, and the company restated its financial statements for two years, with material revenue adjustments totaling $31.7 million.

Case B presented a healthcare company in which revenue anomaly detection identified a persistent divergence between revenue growth (23% annually) and cash flow growth (4% annually) over 12 quarters, with a CFDI of 0.31. The related party component identified limited undisclosed relationship candidates (composite score 0.34), but the revenue anomalies alone produced an elevated composite score of 0.62. Investigation revealed channel-stuffing arrangements with distributors that provided extended payment terms, inflating reported revenues. The company subsequently restated revenue by $89.4 million across six quarters.

Case C demonstrated limitations in the framework for detecting sophisticated international fraud schemes. A manufacturing company with a composite score of 0.28 (Standard Risk) subsequently faced SEC enforcement for undisclosed related-party transactions with offshore suppliers. The relationships involved nominee structures in jurisdictions lacking beneficial ownership transparency, preventing entity matching algorithms from identifying the connections.

Framework validation against 43 SEC enforcement actions involving related-party or revenue fraud showed that the framework correctly identified 34 cases (79.1%) as High or Critical Risk based on pre-enforcement filings, with an additional 6 cases (14.0%) classified as Elevated Risk. Analysis of three undetected cases revealed limitations regarding international subsidiaries and sophisticated round-trip schemes.

5. Conclusion and Implications

5.1. Summary of Key Findings

The rule-based analytical framework demonstrates meaningful capability to identify undisclosed related-party relationships and revenue recognition irregularities. Network analysis identified nearly 2,000 potentially undisclosed relationships across 847 companies, and validation confirmed high accuracy rates for relationship candidates that meet composite score thresholds. Time-series detection achieved strong discrimination, with anomalies strongly associated with restatement and enforcement rates. Companies exhibiting multiple concurrent indicators showed dramatically elevated adverse outcome rates compared to single-indicator or no-indicator observations.

The composite scoring achieves 78.3% precision in high-risk classifications, identifying approximately four-fifths of confirmed fraud cases as elevated risk or higher. The framework's sensitivity-specificity trade-off reflects deliberate calibration that favors detection over minimizing false positives, recognizing the asymmetric costs of undetected fraud versus the additional audit investigation required to detect it.

The framework provides structured analytical tools enhancing fraud detection without requiring specialized technical expertise. The rule-based approach maintains transparency and interpretability, enabling auditors to understand and explain risk assessments. Reliance on publicly available data sources ensures broad applicability across engagement types.

5.2. Implications for Audit Practice

The framework integrates effectively with established methodologies, providing enhanced input to risk assessment procedures. Framework outputs inform fraud risk discussions, guide journal entry testing selection, and identify relationships requiring corroborating evidence.

The framework reinforces professional skepticism by providing independent evidence regarding fraud risk levels. Quantified indicators help auditors overcome anchoring biases from extended client relationships. The framework particularly benefits less experienced team members lacking pattern recognition capabilities developed through extensive audit experience.

Effective deployment requires training that addresses technical operations and the application of professional judgment. Firms should establish quality control processes ensuring consistent application while permitting industry-specific customization. Periodic calibration using updated outcome data maintains detection effectiveness.

5.3. Limitations and Future Research Directions

Framework effectiveness depends substantially on data quality and completeness in source databases. Inconsistent disclosure practices across companies and jurisdictions create coverage gaps that sophisticated fraudsters may exploit. International operations present particular challenges where beneficial ownership transparency varies significantly across regulatory regimes, as demonstrated by Case C analysis.

Historical outcome data limitations constrain empirical validation, as financial reporting fraud represents a low-base-rate phenomenon with significant detection lag. The five-year observation window may not capture all relevant outcomes, potentially underestimating the true effectiveness of the framework.

Future research should examine the integration of alternative data sources, including news media analysis and supply chain data. The framework architecture accommodates extension to additional fraud categories, including expense manipulation and asset valuation irregularities.

References

1. M. A. Khan, U. N. Khan, A. K. Jamali, and J. Jamshed, "The factors contributing to a corporation's demise: An analysis of Enron," *Journal of Management Practices, Humanities and Social Sciences*, vol. 6, no. 2, pp. 15-21, 2022.
2. Y. Wang, L. Zhang, and H. Liu, "Graph neural networks for financial fraud detection: A review," *Frontiers of Computer Science*, vol. 19, no. 1, p. Article 191305, 2024.
3. J. West, and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security*, vol. 57, pp. 47-66, 2016. doi: 10.1016/j.cose.2015.09.005
4. P. M. Dechow, W. Ge, C. R. Larson, and R. G. Sloan, "Predicting material accounting misstatements," *Contemporary Accounting Research*, vol. 28, no. 1, pp. 17-82, 2011.
5. Financial Accounting Standards Board. (n.d.). Accounting Standards Codification Topic 850: Related party disclosures. Norwalk, CT: Author, "." .
6. International Accounting Standards Board, "IAS 24: Related party disclosures," *London, England: IFRS Foundation*, 2009.
7. Y. Chen, and Z. Wu, "Financial fraud detection using the related-party transaction knowledge graph," *Procedia Computer Science*, vol. 199, pp. 733-740, 2022.
8. W. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: A review of anomaly detection techniques and recent advances," *Expert Systems with Applications*, vol. 193, p. Article 116429, 2022.
9. J. L. Perols, "Financial statement fraud detection: An analysis of statistical and machine learning algorithms," *Auditing: A Journal of Practice & Theory*, vol. 30, no. 2, pp. 19-50, 2011.
10. M. Albashrawi, and L. Motiwalla, "Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015," *Journal of Data Science*, vol. 14, no. 3, pp. 553-569, 2019.

11. S. Motie, and B. Raahemi, "Financial fraud detection using graph neural networks: A systematic review," *Expert Systems with Applications*, vol. 240, p. Article 122156, 2024.
12. M. Riskiyadi, "Using data analytics in financial statement fraud detection and prevention: A systematic review of methods, challenges, and future directions," *Journal of Risk and Financial Management*, vol. 18, no. 11, p. Article 598, 2024.
13. M. Jans, N. Lybaert, and K. Vanhoof, "Business process mining of fraud detection in procurement," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4351-4358, 2010.
14. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559-569, 2011.
15. Y. Zhang, and X. Chen, "Accounting data anomaly detection and prediction based on self-supervised learning," *Frontiers in Applied Mathematics and Statistics*, vol. 11, p. Article 1628652, 2025.