Background: The Hedge Fund Fraud Landscape

Section 2 – AI-Based Detection of Hedge Fund Fraud

Joerg Osterrieder

Zurich University of Applied Sciences (ZHAW)

2025

Outline

1. Fraud Taxonomy Overview (5 types, difficulty 1–5)
2. Performance Fabrication
3. Allocation Fraud
4. Strategy Misrepresentation
5. Market Manipulation
6. Regulatory Fraud
7. Data Ecosystem: Four Layers
8. Return Data and Database Biases
9. Regulatory Filings (EDGAR, Forms ADV/D/13F)
10. Alternative Data
11. Synthetic Data
12. Regulatory Context: United States
13. Regulatory Context: European Union
14. Supervisory Technology (SupTech)
15. Section Summary and Key Takeaways

Fraud Taxonomy Overview

Fraud Type	Difficulty	Key Case	Observable Signals
Performance fabrication	3/5	Madoff (2008)	Serial correlation, Benford violations, implausible Sharpe ratios
Allocation fraud	4/5	Petters (2008)	Cross-account return dispersion, win-rate asymmetry
Strategy misrepresentation	3/5	Platinum Partners (2016)	Style drift, factor exposure shifts, textual inconsistencies
Market manipulation	5/5	SAC Capital (2013)	Order-flow anomalies, network centrality, timing patterns
Regulatory fraud	2/5	Lancer Mgmt. (2003)	Filing inconsistencies, text anomalies, omission detection

- Difficulty scale: 1 (straightforward with available data) to 5 (requires privileged real-time data)
- Ordered from most frequently studied to most difficult to detect

Source: Paper Table 1; Section 2.1

Performance Fabrication (Difficulty 3/5)

Definition

- Deliberate misstatement of investment returns
- Forms: Ponzi schemes, return smoothing, NAV manipulation
- Inflating reported value of illiquid positions (OTC derivatives, distressed debt, private placements)

Paradigmatic Case: Madoff

- Fabricated returns for $\geq$20 years
- **Only 7 losing months** across 14 years
- Nearly perfect 45-degree equity curve
- Sharpe ratio exceeded plausible bounds

Detection Signals

- Serial correlation ρ_1 anomalies
- Benford's law violations
- Distributional "kink" at zero (Bollen & Pool, 2012): $\sim$8% of TASS funds flagged
- Implausibly high Sharpe ratios
- Low return volatility relative to stated strategy

Individual signals are noisy, but their **combination** via ML offers substantially improved discriminatory power

Source: Gregoriou (2009); Markopolos (2010); Bollen & Pool (2012); paper Section 2.1.1

Performance Fabrication: Statistical Red Flags

Getmansky et al. (2004)

- Econometric model: serial correlation from managed pricing of illiquid assets
- Typical $\rho_1 = 0.3$–0.5 for illiquid positions
- **Abnormally high** ρ_1 for funds claiming liquid assets $\Rightarrow$ NAV manipulation signal

Bollen & Pool (2012)

- Distributional discontinuity at zero: excess small positive returns, deficit of small negatives
- Correctly identified $\sim$50% of funds that subsequently faced SEC enforcement

Benford's Law

- Tests leading-digit frequency: $P(d) = \log_{10}(1 + 1/d)$
- Applied retroactively to Madoff: anomalies in 9/10 tests
- Limited power for short histories ($<$ 60 months)
- Defeated by knowledgeable fraudster who engineers digit conformity

Key Insight

- Each test captures *one* dimension
- ML classifiers combine signals for multi-dimensional detection

Source: Nigrini (2012); Getmansky et al. (2004); Bollen & Pool (2012); paper Section 2.1.1

Allocation Fraud (Difficulty 4/5)

Definition

- Systematically directing profitable trades to **favored accounts** (proprietary, co-investment)
- Routing losing trades to **client accounts**
- Cherry-picking: delaying allocation until daily P&L known

SEC Evidence

- Favored accounts: **91%** profitable trade allocations
- Client accounts: only **31%**
- Disparity cannot arise by chance

Detection Challenges

- Requires **trade-level data** (order timestamps, fill assignments)
- Rarely available in public databases
- HFR/TASS report only monthly fund-level returns
- Intra-fund allocation patterns entirely obscured

Possible Approaches

- Cross-account return dispersion analysis
- Win-rate asymmetry statistics
- Network-based adviser–account mapping from filings (largely unexplored)

Source: SEC enforcement data; paper Section 2.1.2

Strategy Misrepresentation (Difficulty 3/5)

Definition

- Actual investment behavior diverges materially from stated strategy without adequate disclosure
- Includes:
  - Undisclosed **style drift** (e.g., equity L/S $\rightarrow$ illiquid credit)
  - Leverage misreporting
  - **AI-washing**: falsely claiming AI-driven decisions

Key Case: Platinum Partners (2016)

Detection Methods

- **Change-point detection** (Patton & Ramadorai, 2015): structural breaks in risk exposures often precede fund failure
- Rolling-window factor regressions $\Rightarrow$ style drift detection
- **NLP on Form ADV**: compare stated strategy descriptions vs. quantitative factor exposures
- Cross-modal consistency checking: text vs. numbers

Challenge: distinguishing intentional misrepresentation from legitimate adaptive portfolio management

Source: Patton & Ramadorai (2015); Fung & Hsieh (2001); paper Section 2.1.3

Market Manipulation (Difficulty 5/5)

Forms

- **Front-running** client orders
- **Insider trading** (incl. "shadow trading" on economically related securities)
- **Spoofing**: large orders placed and rapidly cancelled

Key Case: SAC Capital (2013)

- Guilty plea to insider trading charges
- **$1.8 billion** penalty

Why Most Difficult

- Requires **real-time, tick-level** trade and order-book data
- Not available in standard hedge fund databases
- Detection needs:
  - Network analysis of communication patterns
  - Temporal analysis of orders vs. MNPI events
  - Cross-market surveillance for shadow trading

SEC MIDAS

- Processes $\sim$1 billion records/day
- Academic research constrained by data access

Source: Lewis (2012); SEC MIDAS; paper Section 2.1.4

Regulatory Fraud (Difficulty 2/5)

Definition

- Materially false or misleading information in mandatory filings
- Key filings:
  - **Form ADV**: uniform registration (Investment Advisers Act)
  - **Form D**: Reg D offering notices
  - **Form 13F**: quarterly holdings ($100M+ managers)
- Ranges from deliberate misstatements to material omissions (disciplinary history, conflicts)

Why Most Tractable

- Data are **structured / semi-structured**
- **Publicly accessible** via EDGAR
- Amenable to traditional text analysis *and* modern NLP
- Cross-referencing filings with external data (e.g., AUM vs. implied fund flows)

Key Results

- Dimmock & Gerken (2012): Form ADV predicts SEC enforcement actions
- Brown et al. (2008): filing data contain fraud-relevant signals complementing return-based tests

Source: Dimmock & Gerken (2012); Brown et al. (2008); paper Section 2.1.5

Fraud Taxonomy: Visual Summary

: 01_fraud_taxonomy – Radar or matrix chart showing five fraud types with detection difficulty, data requirements, and AI method suita

Source: Paper Table 1; Section 2.1

Data Ecosystem: Four Layers

Layer 1: Return Data

- Lipper TASS, HFR, BarclayHedge, Morningstar
- Monthly return series, self-reported characteristics
- 7,000+ live and defunct funds in TASS

Layer 2: Regulatory Filings

- SEC EDGAR: Forms ADV, D, 13F
- Structured (XML) + free-text (PDF)
- Post-Dodd-Frank: $150M+ must register

Layer 3: Alternative Data

- News/social media sentiment, satellite imagery, web traffic, litigation records
- Market: **$7.5B** (2023), projected $273B by 2032

Layer 4: Synthetic Data

- Addresses extreme class imbalance
- SMOTE, GANs, VAEs
- Privacy-preserving generation for cross-institutional collaboration

Detection effectiveness is bounded by data quality, coverage, and granularity

Source: Paper Section 2.2

Return Data: Three Critical Biases

Survivorship Bias

- Defunct funds exit live databases
- Overstates average returns by $\sim$**+242 bp/year**
- Fung & Hsieh (2009)

Backfill Bias

- Retroactive submission of favorable pre-reporting history
- Overstates returns by $\sim$**+442 bp/year**
- Also called "instant history bias"

Selection Bias

- Voluntary reporting
- Strong track records more likely to report
- Distressed/fraudulent funds may stop reporting before detection

Pernicious asymmetry for fraud detection: detected frauds enter graveyard; undetected frauds remain in live data, contaminating the "clean" training class.

Source: Fung & Hsieh (2009); Agarwal et al. (2011); paper Section 2.2.1

Regulatory Filings (EDGAR)

Key Filing Types

- **Form ADV Part 1**: structured XML via IARD
  - Business, ownership, clients, disciplinary history
- **Form ADV Part 2** ("brochure"): free-text PDF
  - Strategies, fees, risk factors, conflicts
  - No standardized structure
- **Form 13F**: quarterly equity holdings ($100M+ managers)
  - Machine-readable but contains known errors

Practical Challenges

- Merging data across filing types
- Linking to commercial return databases (TASS $\leftrightarrow$ SEC CRD numbers)
- Substantial data engineering effort

Value for Fraud Detection

- Cross-validation: return-based flags vs. filing-derived signals
- Auditor changes, custody arrangements, disciplinary histories
- Post-Dodd-Frank: dramatically expanded disclosure universe

Source: Brown et al. (2008); Dimmock & Gerken (2012); paper Section 2.2.2

Alternative Data

Sources

- News and social media sentiment
- Satellite imagery and geolocation data
- Web traffic analytics
- Patent filings and litigation records

Market Size

- $\sim$**\$7.5 billion** (2023)
- Projected **\$273 billion** by 2032
- Driven by adoption among investment managers and regulators

Fraud Detection Uses

- Flag reputational signals *before* regulatory actions
- Independently verify economic claims (e.g., satellite foot traffic vs. reported performance)
- Triangulate plausibility of stated strategies/returns

Risks

- Sentiment signals are noisy, susceptible to manipulation
- Acquisition/processing costs are substantial
- Privacy concerns (intersect with EU AI Act)
- Indirect relationship to fund-level fraud signals

Source: Paper Section 2.2.3

Synthetic Data for Class Imbalance

The Problem

- Extreme class imbalance: confirmed frauds = small fraction of total
- 50–100 confirmed cases vs. 10,000+ funds
- Difficult to train supervised classifiers

Methods

- **SMOTE**: interpolation between existing positive examples (most widely used)
- **GANs**: learn distributional properties of known fraudulent funds
- **VAEs**: probabilistic generation with calibrated uncertainty

Validation Challenges

- Must preserve statistical dependencies and temporal dynamics
- Cannot amplify artifacts of training data
- Generative models > interpolation for realism and diversity

Privacy-Preserving Generation

- Regulators cannot share enforcement-labelled data
- Differentially private generative models could enable synthetic dataset release
- Preserves aggregate statistics while protecting identities
- Remains largely aspirational in hedge fund domain

Source: Chawla et al. (2002); paper Section 2.2.4

Data Ecosystem: Visual Summary

ur-layer data architecture diagram showing Return Data, Regulatory Filings, Alternative Data, and Synthetic Data with coverage, frequ

Source: Paper Section 2.2

Dodd-Frank Act (2010)

- Eliminated "private adviser exemption"
- Advisers with AUM $\geq$\$150M must register with SEC
- Dramatically expanded universe of funds subject to disclosure
- Created filing data underpinning many detection approaches

SEC Divisions

- **DERA** (est. 2009): quantitative analysis for enforcement & rulemaking
- **MIDAS** (since 2013): $\sim$1B records/day from all equity exchanges
- **CRQA**: trading pattern databases from past enforcement, prioritizes future exams

Whistleblower Program

- Established under Dodd-Frank §922
- Response to failure to act on Markopolos warnings
- Over **\$1.5 billion** awarded to whistleblowers
- Thousands of tips complementing algorithmic detection

Hybrid Detection Paradigm

- Human intelligence (tips) + machine intelligence (quantitative screening)
- Underexplored synergy between these two channels

Source: Brown et al. (2008); SEC (2023); paper Section 2.3.1

Regulatory Context: European Union

AIFMD (2011)

- Harmonized framework for alternative investment fund managers
- Reporting obligations, leverage limits, investor disclosure
- Structured data analogous to US Form ADV regime

EU AI Act (Regulation 2024/1689)

- Entered into force August 2024
- AI for fraud detection classified as **high-risk**
- Mandatory requirements:
  - Risk management systems
  - Data governance standards
  - Technical documentation
  - Human oversight
  - Transparency obligations (Art. 13)

Implications for Model Selection

- Art. 13: outputs must be "sufficiently transparent to enable deployers to interpret"
- Opaque models (deep neural networks) may require post-hoc explainability (SHAP, LIME)
- Interpretable models (logistic regression, decision trees) may be preferred despite potentially lower performance

Accuracy vs. Explainability Tension

- Detection system that cannot articulate basis for suspicion $\Rightarrow$ limited regulatory use
- EU AI Act codifies this intuition into law
- Similar requirements likely to emerge globally

Source: EU AI Act (2024); paper Section 2.3.2

Supervisory Technology (SupTech)

Definition

- Advanced analytics and AI used by **financial regulators and central banks** to enhance oversight
- Shift from *reactive enforcement* (investigate after losses) to *proactive surveillance* (detect anomalies before escalation)

Adopters

- SEC (DERA, CRQA, MIDAS)
- Bank of England
- Monetary Authority of Singapore
- De Nederlandsche Bank

Approach

- Establish baseline behavioral profiles for regulated entities
- Flag deviations exceeding statistical thresholds
- Conceptually similar to Bollen & Pool (2012) but at institutional scale

Challenges

- False positives consume scarce examination resources
- Sophisticated fraudsters may reverse-engineer detection criteria
- Shortage of staff with combined regulatory + ML expertise
- **Cross-border coordination** limited: Cayman Islands registration, NY management, EU investors, Asian venues

Source: FSB (2017); BIS (2024); paper Section 2.3.3

Regulatory Landscape: Comparison

scape – Side-by-side comparison of US (Dodd-Frank, SEC/DERA/MIDAS) vs. EU (AIFMD, EU AI Act) regulatory frameworks with Su

Source: Paper Section 2.3

Section Summary

Fraud Taxonomy

- 5 types: performance fabrication, allocation fraud, strategy misrepresentation, market manipulation, regulatory fraud
- Difficulty ranges from 2/5 (regulatory) to 5/5 (market manipulation)
- Each type requires different data and detection approaches

Data Ecosystem

- 4 layers: return data, regulatory filings, alternative data, synthetic data
- Severe biases: survivorship (+242 bp), backfill (+442 bp), selection

Regulatory Context

- US: Dodd-Frank expanded disclosure; SEC investing in computational enforcement
- EU: AIFMD for fund oversight; EU AI Act imposes high-risk requirements on detection AI
- SupTech: proactive surveillance emerging but faces cross-border and capacity challenges

These foundations establish what must be detected, which data are available, and what constraints govern AI systems.

Source: Paper Section 2

Key Takeaways

1. Hedge fund fraud spans **five distinct types** with fundamentally different data requirements and detection difficulty
2. The data ecosystem is **rich but biased** – survivorship, backfill, and selection biases directly affect model training
3. Post-Dodd-Frank regulatory filings create a **valuable structured data source** that did not exist before 2012
4. The **alternative data revolution** ($7.5B market) opens new detection avenues but introduces noise and privacy concerns
5. **Extreme class imbalance** (50–100 confirmed cases) necessitates synthetic data generation
6. The EU AI Act imposes **mandatory transparency requirements** on high-risk AI fraud detection systems
7. SupTech represents a paradigm shift from reactive to **proactive surveillance**, but cross-border coordination remains limited
8. **No single data layer or detection method** is sufficient – multi-modal AI integration is essential

Source: Paper Section 2