



U.S. Securities and Exchange Commission

The Hope and Limitations of Machine Learning in Market Risk Assessment

Scott W. Bauguess

*Deputy Chief Economist
U.S. Securities and Exchange Commission*

March 6, 2015

Disclaimer

The Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The views expressed herein are those of the authors, and do not necessarily reflect the views of the Commission or of the authors' colleagues on the staff of the Commission.



What is Machine Learning?

Google says

- The scientific discipline that explores the construction and study of algorithms that can learn from data. ([wikipedia.com](https://en.wikipedia.org))
- The science of getting computers to act without being explicitly programmed. ([coursera.org](https://www.coursera.org))
- A type of artificial intelligence (AI) that provides computers with the ability to learn...similar to that of data mining. ([whatis.com](https://www.whatis.com))
- ML solves problems that cannot be solved by numerical means alone ([toptal.com](https://www.toptal.com))
 - *Supervised: train on a predefined set of examples*
 - *Unsupervised: give program a bunch of data to find patterns*
- A computer's way of learning from examples ([businessinsider.com](https://www.businessinsider.com))

World's first machine learning program by a pioneer in AI research:

Arthur Samuel, 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal* **3** (3): 210–229.



SEC Risk Assessment Activities

SEC Mission

- Protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation

Role of Risk Assessment

- Identification of Misconduct
 - *Statutory and rule violations and fraud*
- Disclosure review
 - *Forms and filings made by SEC registrants*
 - *Facilitate market evaluation of investment risk*
- Evaluation of Market-wide Risks
 - *Financial stability (FSOC, FSB)*



SEC Risk Assessment Activities

Types of Entities and Individuals of interest

- Asset Managers
 - *Funds registered under 1940 Investment Company Act*
 - *Investment advisors registered under 1940 Investment Advisers Act*
 - *Private funds (e.g., hedge funds)*
- Public Companies
 - *Governed by 1933 Securities Act & 1934 Securities Exchange Act*
 - *Issuers of securities that generally trade on national market exchange*
 - *Periodic and event driven disclosure on forms 10-K, 10Q, 8-K*
- Private Companies
 - *Issue securities that can trade over-the-counter (OTC)*
 - *Minimal disclosure requirements*



Data/Risk Analytics at the Commission

Division of Economic and Risk Analysis

- Office of Risk Assessment
- Office of Research and Data Services

Office of Compliance Inspections and Examinations

- Office of Risk Analysis and Surveillance
- Quantitative Analytics Unit

Division of Enforcement

- Center for Risk and Quantitative Analytics

Division of Trading and Markets

- Office of Analytics and Research

Division of Investment Management

- Risk and Examinations Office



ML applications at Commission

Text Analytics

- MD&A discussion – identify herding behavior in disclosures
- Classification of content in unstructured filings

Identify illicit trading activity

- Transaction data is generally well behaved and outcomes (e.g., rule violations) measureable
- ENF – Center for Risk and Quantitative Analytics
 - *Developing analytical tools that use pattern analysis to more easily identify suspicious activity and find connections among traders.*
- OCIE – Quantitative Analytics Unit
 - *National Exam Analytics Tool (NEAT) allows onsite examiners to download millions of transactions at an Investment Adviser to quickly identify signs of front running, cherry picking, window dressing, etc.*



ML Challenges in Risk Assessment

Need observable actions (**behaviors**) and outcomes (**rule violations**)

- Checkers: series of **moves** that ends in **win/loss**
- Online retailing: shoppers who **bought this** also **bought that**
- Investor: series of **trades** and potential **use of inside information**
- Hedge fund: series of **self-reported returns** and the **misvalued assets**
- Public company: series of **disclosures** and **material misstatement**

Misconduct is a latent variable (unless detected)

- Need to generate signals that are a proxy for illicit behavior
- Build models that establish correlations between observable actions and instances of detected misconduct



ML Challenges in Risk Assessment

Omitted factors and spurious correlations

- Correlation may not be causation

False positives

- Detection of average v. specific effects

Operationalizing a successful proof of concept

- A tool can't be more sophisticated than the user of it

Over training models

- How do you incorporate new (illicit) behavior?

Paradox of a successful model

- Successful model should “prevent” misconduct
- Less misconduct may make it more difficult to calibrate the model



Measuring Signals of Risk

Market-based signals

- Securities prices – incorporate lots of signals; hard to disentangle
- News articles – qualitative

Disclosure items in regulatory filings

- SEC forms and filings (numeric and narrative)
- Forms/questions designed with causality in mind (e.g., a survey)
- Answers subject to bias (liars lie)
- Information may not be structured
 - *Narrative disclosures (management discussion and analysis)*
- Numeric data may be structured, but not standardized
 - *Can't compare across market entities*

Tips, Complaints, and Referrals



Detecting Misconduct

Estimate a model based on historical data

- Identify outcomes (LHS variables)
 - *Instances of misconduct (e.g., enforcement actions)*
 - *Risky behavior that may be indicative of misconduct*
- Identify risk factors (RHS variables)
 - *E.g., entity performance, intrinsic characteristics, and actions taken*
- Estimate model using out-of-sample data
- Apply factor loadings to new data as it is collected
 - *Generate predictions and use to risk rank*
 - *Examine changes to realized risk factors that contribute to the prediction*
- Recalibrate the model
 - *Use new outcomes from application of model*



Development Lifecycle

I. Understand market: Have experts

II. Develop quantitative measures of actions and outcomes:
data from forms, filings, news, and other market sources

III. Build a model: Estimate using historical data

IV. Apply output and assess outcomes

V. Generate feedback: educate experts, update required disclosure, recalibrate model

Hedge fund risk assessment

Aberrational Performance Inquiry (API) Model

- The SEC Enforcement Division's Asset Management Unit uses proprietary risk analytics to evaluate hedge fund returns.
- Performance that appears inconsistent with a fund's investment strategy or other benchmarks forms a basis for further scrutiny

Methodology

- Use hedge-fund self-reported returns (vendor purchased)
- Classify hedge-funds according to strategy (peer funds)
- Analyze strategy-adjusted returns in a number of dimensions to assess the likelihood they accurately reflect the underlying assets

Early quantitative risk assessment success

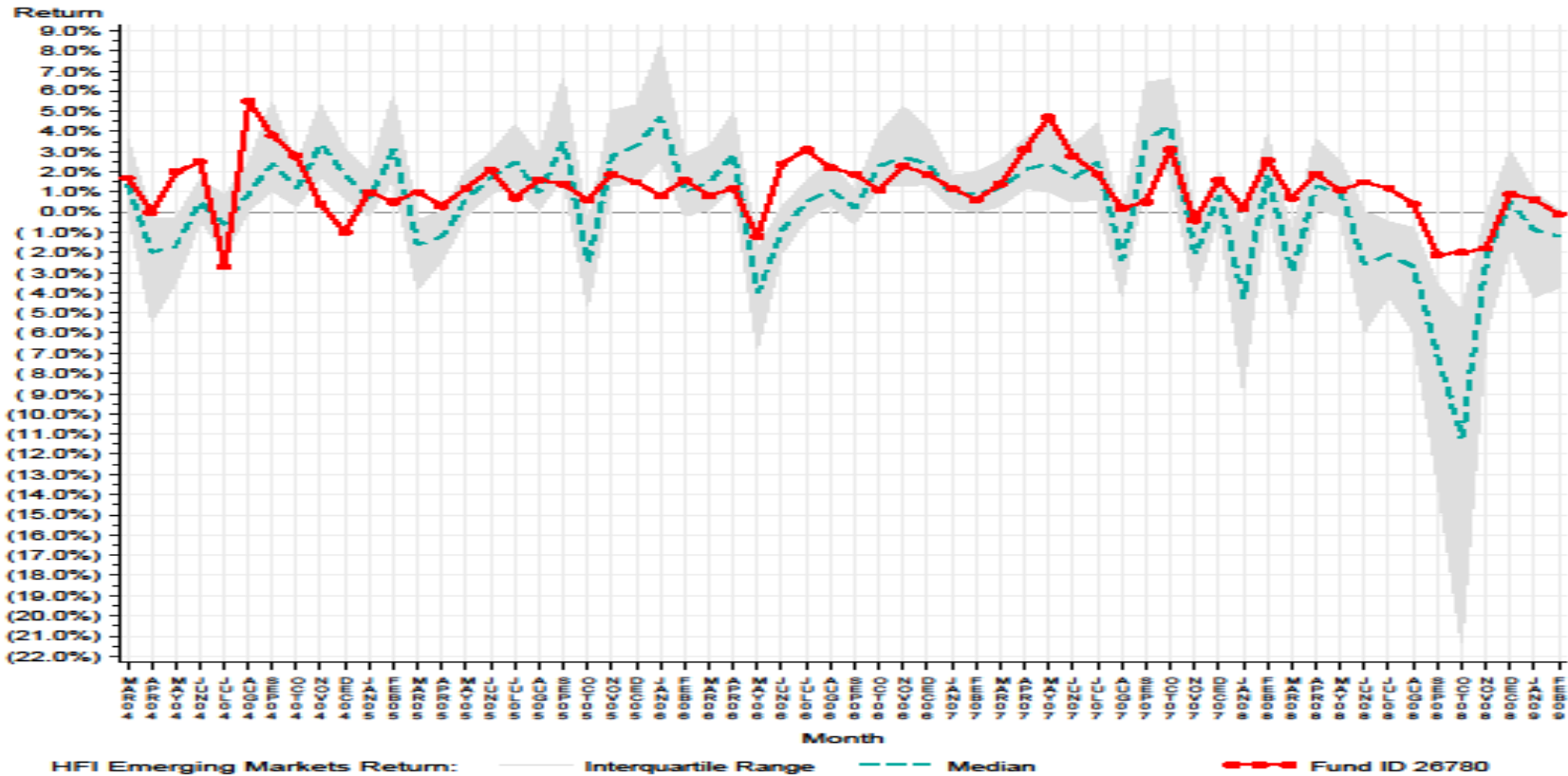
- A number of successful cases since 2009 that were identified using this tool



Hedge fund risk assessment

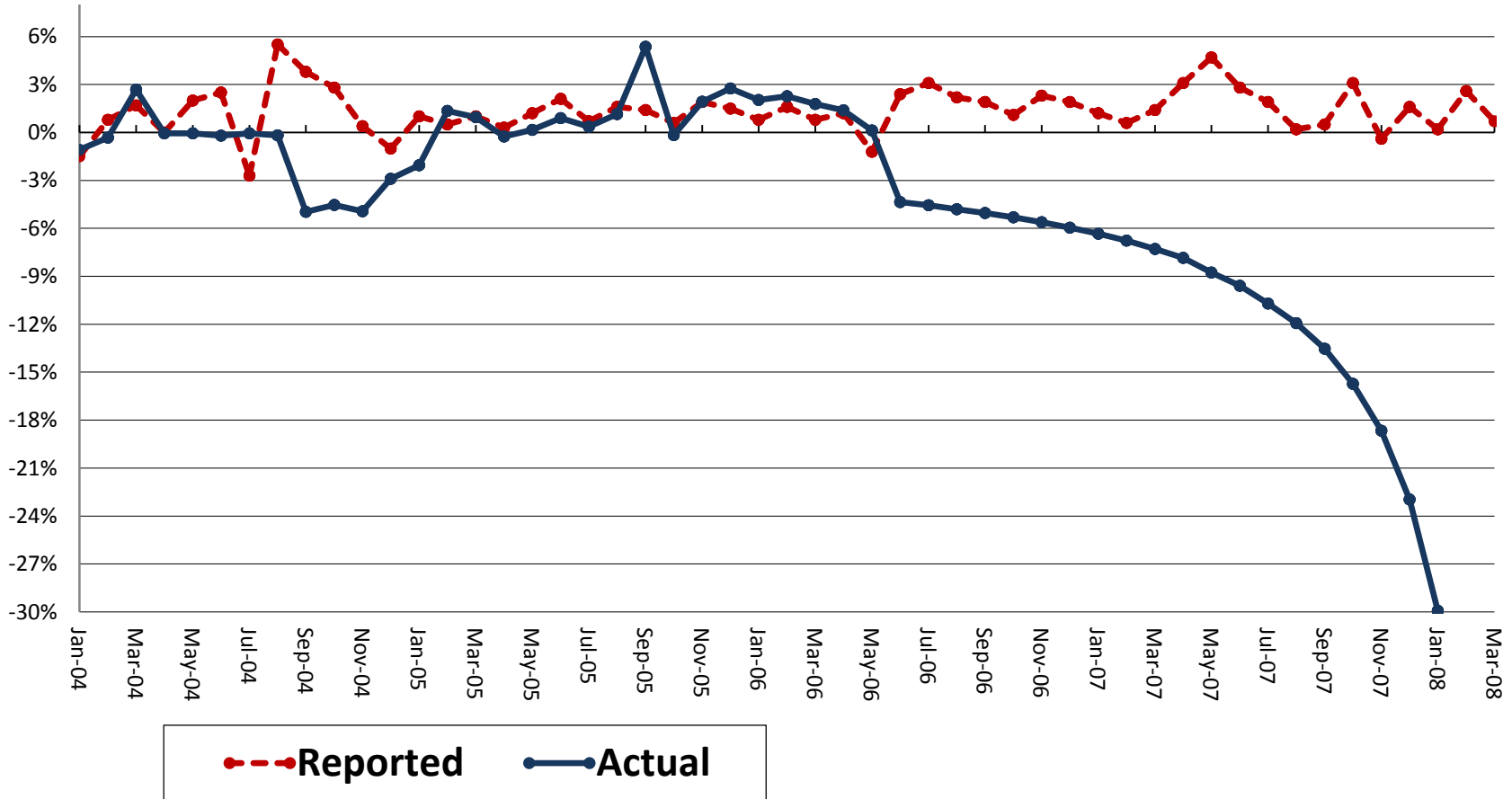
Reported returns v. strategy

Figure 2. Monthly Returns over Time (HFI Peer Group vs Fund ID 26780)



Hedge fund risk assessment

The Reality: Actual vs. Reported Returns



API Outcomes

In re: GLG Partners, Inc. (Dec. 12, 2013)

SEC v. Yorkville Advisors (S.D.N.Y. Oct. 17, 2012)

SEC v. Balboa (S.D.N.Y. Dec. 1, 2011)

SEC v. Rooney (N.D. Ill. Nov. 18, 2011)

In re: LeadDog Capital Markets, LLC (Nov. 15, 2011)

SEC v. Kapur (S.D.N.Y. Nov. 10, 2011)

SEC v. Reid (S.D. Ga. Feb. 1, 2011)

SEC v. Jeffrey R. Neufeld (N.D. Ill. Apr. 19, 2010)



Corporate Issuer Risk Assessment (CIRA)

Dashboard tool developed for disclosure review and investigative staff

- Incorporates (modeled) measures of earnings management
- Standard financial ratios
- Key financial metrics

Allows comparison to peer firms

- Custom, Geographic or industry-specific

Users can generate risk rankings based on Company behaviors

- Risky behaviors identified by determinants of misconduct derived from previous enforcement actions.
- E.g., a company with poor stock price performance, high inventory growth, low sales growth, and a change of auditor may be indicative of aggressive earnings management.



Corporate Issuer Risk Assessment

Discretionary accruals and earnings management

- A regression model for *non-discretionary* accruals is specified that is a linear function of financial statement variables
- All regress a measure of accruals (total or working capital) on factors that proxy for non-discretionary components. The *residual* is treated as the estimate of discretionary accruals

Standard accounting model (Modified-Jones)

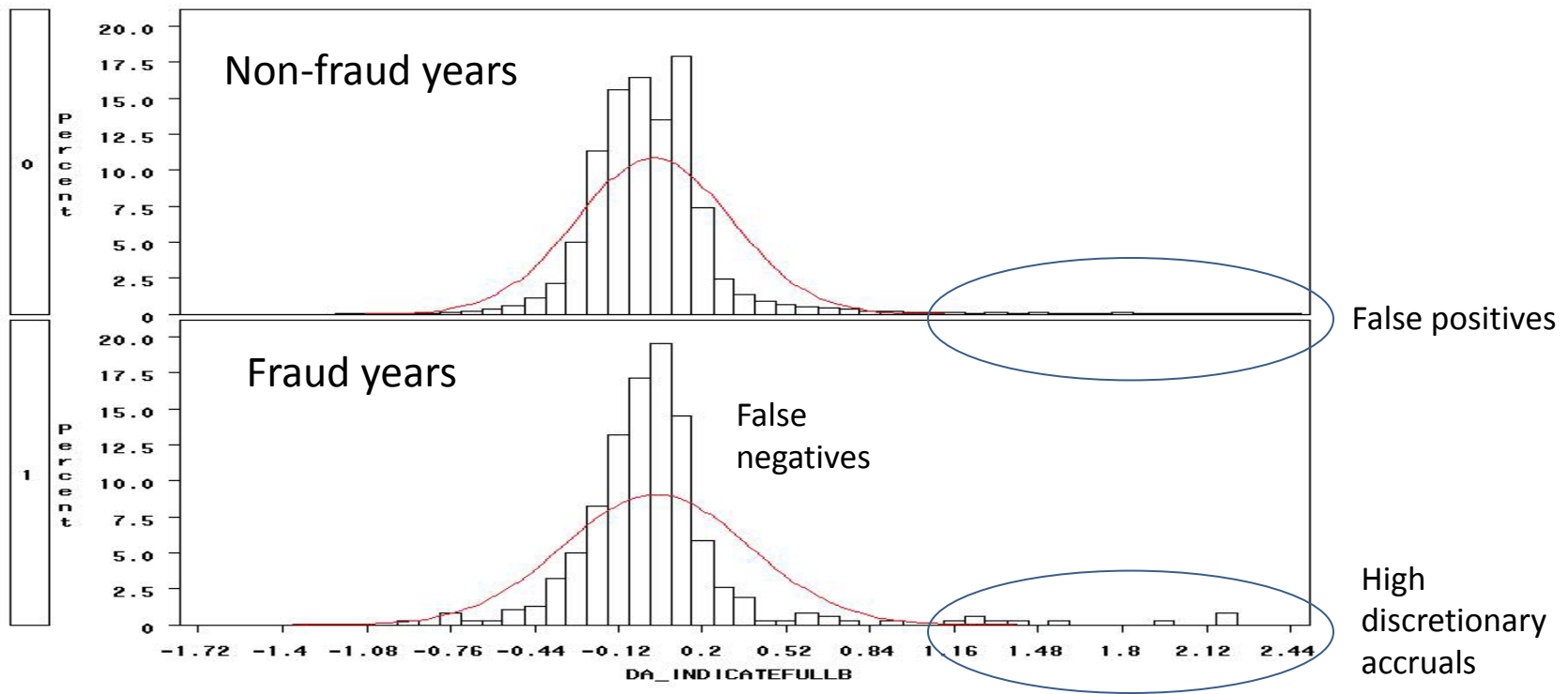
$$\frac{TA}{Assets} = \alpha_0 \frac{1}{Assets} + \alpha_1 \frac{\Delta(REV-REC)}{Assets} + \alpha_2 \frac{PPE}{Assets} + \epsilon$$

- Residual is interpreted as “discretionary accruals” and it could be symptomatic of earnings management



Corporate Issuer Risk Assessment

Establish correlation between discretionary accruals and instances of Fraud or misconduct.



CIRA Dashboard

-- sample company output --

Dashboard highlights the standard financial ratios for each issuer that are outliers within the issuer's industry or when compared to all other issuers during analysis period

<input type="radio"/> Industry <input checked="" type="radio"/> All				<input checked="" type="radio"/> 2013 <input type="radio"/> 2012 <input type="radio"/> 2011 <input type="radio"/> 2010		49 Electric, Gas, and Sanitary Services		
RATIOS								
Liquidity	Leverage	ROA & ROE	Margin	Turnover	Days	Valuation Ratios	A/R & Revenues	Other
Current ratio	Short-term debt to total assets	ROA	Gross profit margin	Asset turnover	Days working capital	Market-to-book ratio	Annual revenue growth rate	Cash flow index
Quick ratio	Long-term debt to total assets	ROE	EBITDA profit margin	Fixed asset turnover	Days sales outstanding	Price-to-earnings ratio	3 year revenue compound annual growth rate	FF&E to total assets
Interest coverage ratio	Total debt to total assets		Operating expense ratio	Inventory turnover	Days inventory	Price-to-sales ratio	Annual A/R growth rate	XBRL custom tag - percentage
	Total debt to total stockholders' equity		Net profit margin	Accounts receivable turnover	Days payables outstanding		3 year A/R compound annual growth rate	
	Interest expense to total debt (interest rate)						1 year A/R growth rate minus revenue growth rate	
							3 year A/R growth rate minus revenue growth rate	
							A/R to sales	

Select a ratio above to add to the dashboard.



Successful Risk Assessment

Good data

- Structure can matter more than amount of data
- Incorporate results into (modified) rules and required disclosures

Well trained “humans”

- Ability to understand the risk factors and apply the models
- Tolerance for false positive

Collaboration among many skill sets

- Model builders (economists/engineers)
- IT specialists to implement tools
- Institutional knowledge of how markets actually work
- Data scientists (know how data actually works)

