Can Textual Analysis of Corporate Filings Predict Business Related Risks? A look at Analytic Geometry in Textual Analysis

Dr. Rajendra P. Srivastava

PhD (Physics, 1972), PhD (Accounting, 1982)

Professor Emeritus, Ex EY Professor, and

Ex Director of EY CARAT

The University of Kansas, USA

&

Founder & CEO, SeekEdgar, LLC

Outline

- Textual Analysis
- Analytical Geometry in Textual Analysis
- Risk Assessment Models based on Textual Analysis
- Demo of SeekiNF Technology
- Conclusion
- Questions & Answers

Textual Analysis

- Word Counts, Sentence Counts, and Phrase Counts
- Readability Indices (Difficult to Read and Comprehend)
- Tone of a Document (Positive, Negative)
- Tone Dispersion (Concentrated or Dispersed)
- Word Variation from Year to Year
- Risk Assessment Models
  - Risk in General
  - Financial Risk, Litigation Risk, Tax Risk
  - Idiosyncratic Risk & Systemic Risk
  - Fraud Risk, etc.
- Market Competition using Proximity Counter

Since I am neither a salesperson nor a businessman, top Universities are helping me spread the word. For example, Professor Mike Minnis of the University of Chicago introduced me to several Schools including HEC-Paris (to Professor Pape Kraft at HEC). I was scheduled to visit HEC last year but because of COVID I had to cancel my visit. Here is an excerpt.

Hi Pepa,
Hope all is well. Just wanted to make a quick introduction to Raj Srivastava, an accounting academic who started the company "SeekEdgar" several years ago. . . . . I thought I'd connect the two of you in case you have any interest. I use SeekEdgar for both research and teaching—it is an excellent textual analysis tool for all sorts of SEC document. Very cool tool and I have known Raj for several years now. . . . . .

Cheers!
Mike

Another example, Professor Bob Herz (former Chairman of FASB), Columbia School of Business, once attended my presentation at Rutgers University, and was very impressed with the technology. He wrote to several schools including INSEAD in France. INSEAD invited me to make a presentation and were happy to host me as a professor. Unfortunately, again, I had to cancel it because of COVID. Here is a copy of his email.

Hi Sharon. It's Bob Herz. I hope all is going well for you at Insead. I recently attended a conference on new accounting and auding technologies at Rutgers Business School. Among the very interesting presentations was one by Dr. Rajendra Srivastava, CEO of a company called SeekEdgar, on text mining and textual analysis of SEC filings, PCAOB reports, and millions of other documents. His presentation included a number of very interesting applications that speed and enhance research and were of great interest to the accounting, audit, and finance academics and others in the audience (and which I believe would also be of interest to professional investors such as hedge funds). Many leading universities are customers. I recently introduced him to the accounting department at Columbia Business School and I thought it may be of interest to you and your colleagues at Insead. Dr. Srivastava would like to make a presentation on the SeekEdgar Technology to the relevant people at Insead. I am copying Dr. Srivastava on this note and leave it up to the two of you to connect. Many thanks.

Textual Analysis:
Seven Readability Indices

SeekEdgar

Readability Indices

1. Gunning-Fog Index https://en.wikipedia.org/wiki/Gunning_fog_index
2. Smog Index https://en.wikipedia.org/wiki/SMOG
3. Flesch Reading Ease https://en.wikipedia.org/wiki/Flesch–Kincaid_readability_tests
4. Flesch-Kincaid Grade Level https://en.wikipedia.org/wiki/Flesch–Kincaid_readability_tests
5. Automated Readability Index https://en.wikipedia.org/wiki/Automated_readability_index
6. Coleman-Liau Index https://en.wikipedia.org/wiki/Coleman–Liau_index
7. Bog Index https://kelley.iu.edu/bpm/activities/bogindex.html

1. Gunning-Fog Index

(Robert Gunning, 1952)

Gunning-Fog Index = 0.4[(Words/Sentences)
+ 100(Complex words/Words)]

- 17 College graduate
- 16 College senior
- - - - -
- 12 High school senior
- - - - -
- 10 High school sophomore
- - - - -
- 6 Sixth grade

2. Smog Index

$$\text{Grade} = 1.043\sqrt{\text{number of polysyllables}\frac{30}{\text{number of sentences}}} + 3.1291$$

3. Flesch Reading Ease

(Technical Materials, developed under a contract to the U.S. Navy in 1975)

Flesch reading ease

= 206.835 -1.015(total words/total sentences)

- 84.6(total syllables/total words)

Scores

- 100.00-90.00, 5th grade.
- 90.0–80.0, 6th grade
- - - - -
- 30.0–0.0, College Graduate - Very difficult to read.

4. Flesch-Kincaid Grade Level

(extensively used in the field of education)

Flesch-Kincaid Grade Level

= 0.39(total words/total sentences)

+ 11.8(total syllables/total words)

- 15.59

Scores relate to the US Grade levels

5. Automated Readability Index

Automated Readability Index (ARI)

= 4.71(characters/words)

+ 0.5(words/sentences) – 21.43

where characters is the number of letters and numbers, words is the number of spaces, and sentences is the number of sentences.

6. Coleman-Liau Index

Coleman-Liau Index = 0.0588**L** -0.296**S** – 15.8

L is the average number of letters per 100 words and **S** is the average number of sentences per 100 words, i.e.,

L = 100(Letters/Words)

S = 100(Sentences/Words)

7. Bog Index

A plain English measure of financial reporting readability

by

Bonsall IV, Leone, Rennekamp

in

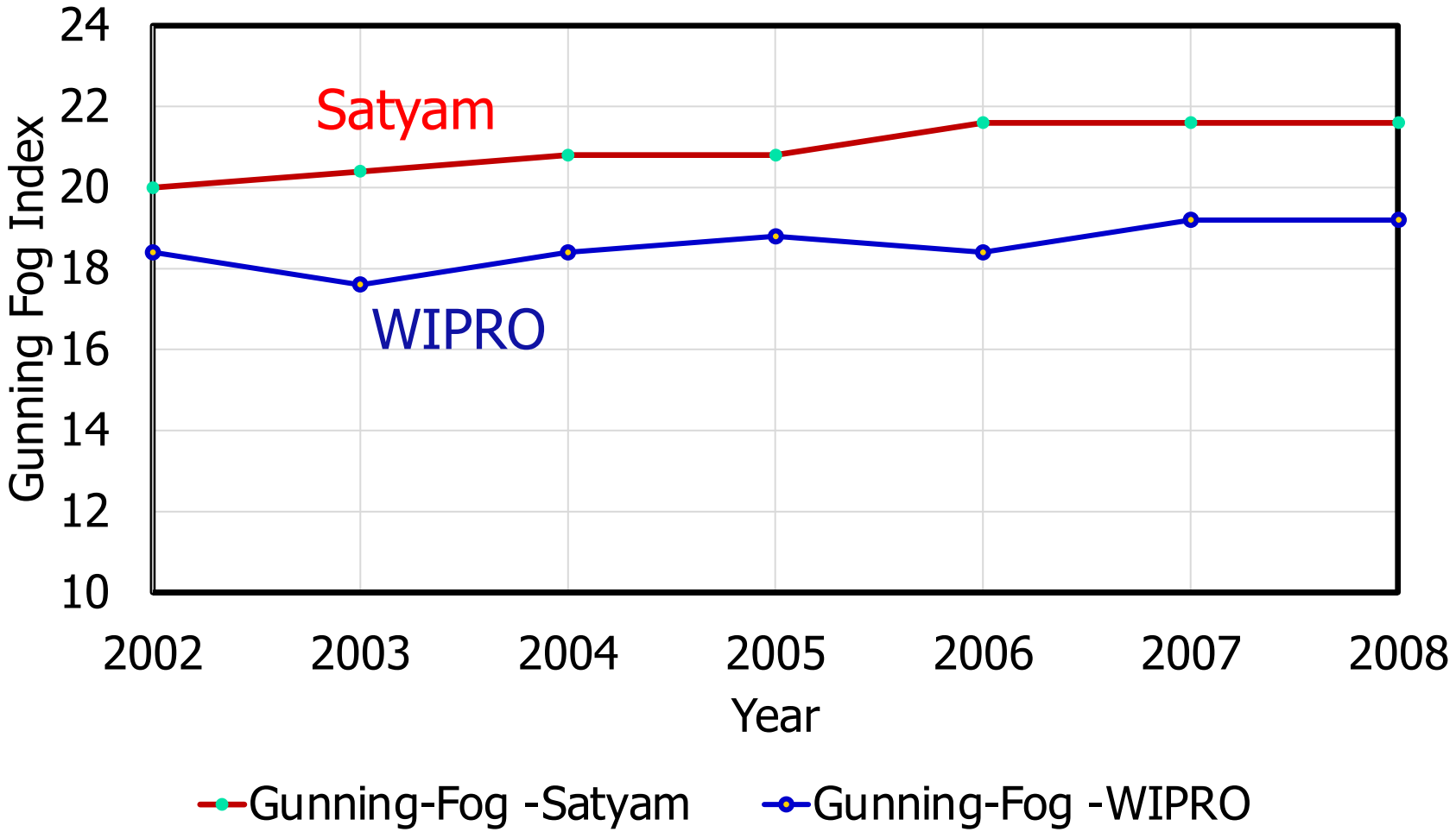
Journal of Accounting and Economics, 63 (2017) pp. 329–357

Example: Readability Indices for Satyam and WIPRO

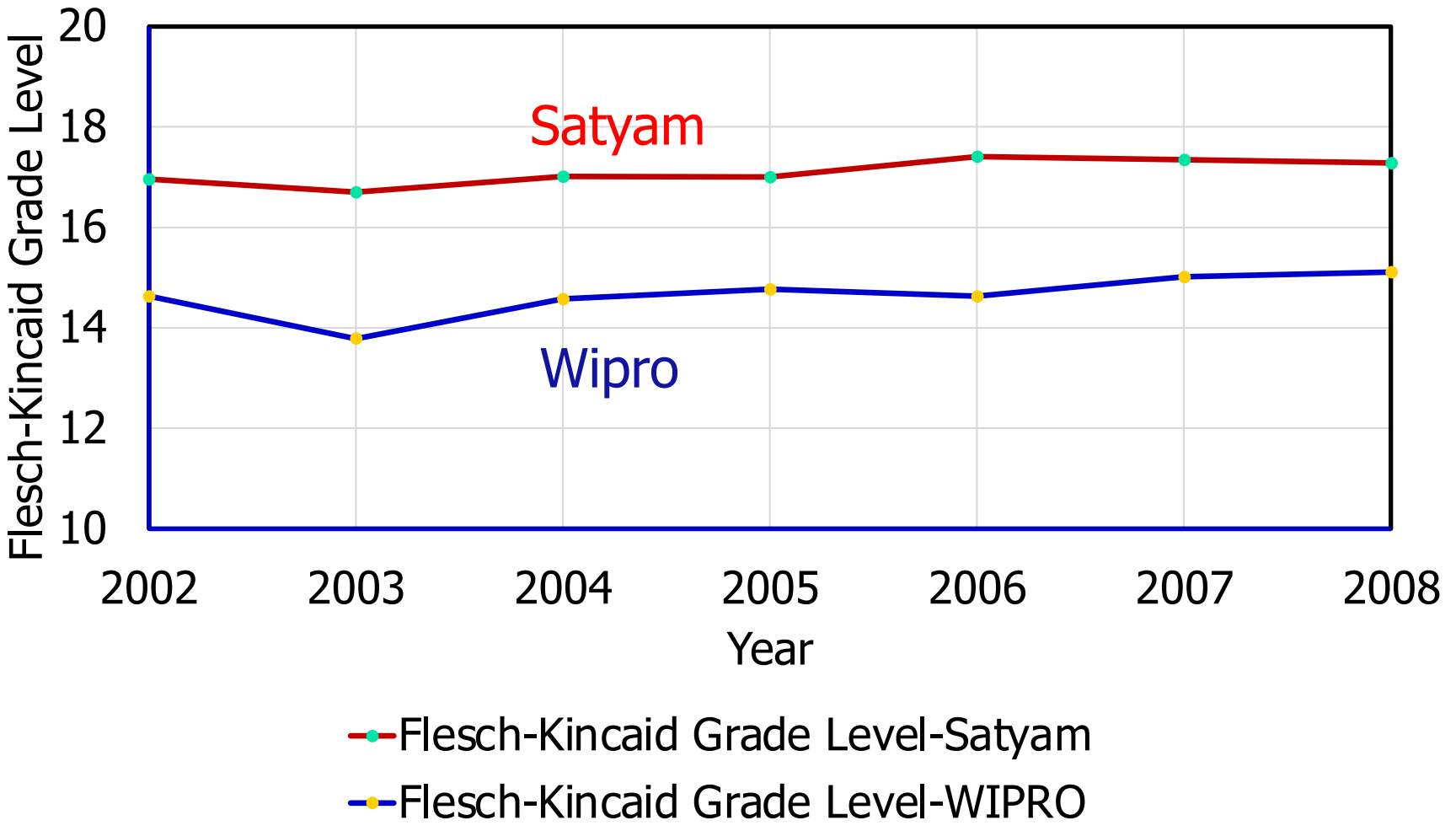
Satyam Textual Analysisç							
Year	2008	2007	2006	2005	2004	2003	2002
Total Word Count	81258	85673	80785	58473	67858	70837	259828
Total Word Count without numerics	74833	79145	74881	54641	60675	63526	227833
Sentence Count	2642	2770	2575	1966	2175	2368	5770
Gunning-Fog Index	21.6	21.6	21.6	20.8	20.8	20.4	20
Smog Index	18.666	18.762	18.73	18.459	18.394	18.18	13.618
Flesch Reading Ease	21.777	21.777	22.212	22.893	22.92	23.236	51.699
Flesch-Kincaid Grade Level	17.281	17.344	17.411	17.001	17.014	16.704	16.962
Automated Readability Index	17.759	17.819	17.908	17.316	17.383	16.964	13.404
Coleman-Liau Index	14.439	14.357	14.145	14.2	14.239	14.386	0.293

WIPRO LTD							
Year	2008	2007	2006	2005	2004	2003	2002
Total Word Count	93966	99464	96763	101922	87781	75005	120396
Total Word Count without numerics	85584	90570	88177	93798	78915	66793	104844
Sentence Count	3624	3894	3865	4080	3511	3290	4656
Gunning-Fog Index	19.2	19.2	18.4	18.8	18.4	17.6	18.4
Smog Index	17.059	16.935	16.644	16.797	16.625	16.004	16.688
Flesch Reading Ease	28.928	28.956	30.938	30.24	30.684	32.505	30.451
Flesch-Kincaid Grade Level	15.113	15.017	14.628	14.771	14.579	13.785	14.629
Automated Readability Index	15.082	15.012	14.573	14.709	14.397	13.423	14.471
Coleman-Liau Index	14.04	14.182	13.916	13.97	13.9	14.05	13.959

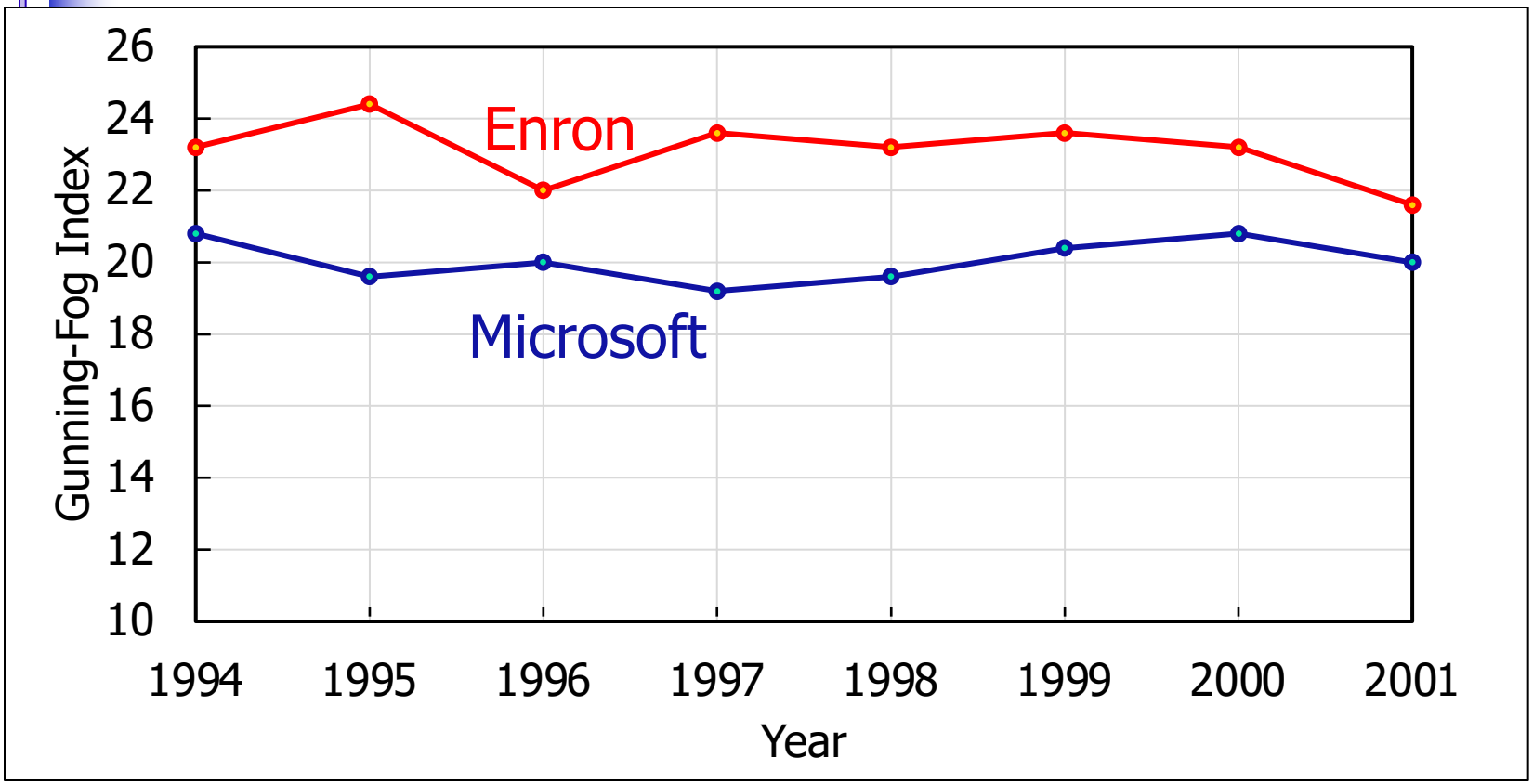
Gunning Fog Index for Satyam and WIPRO



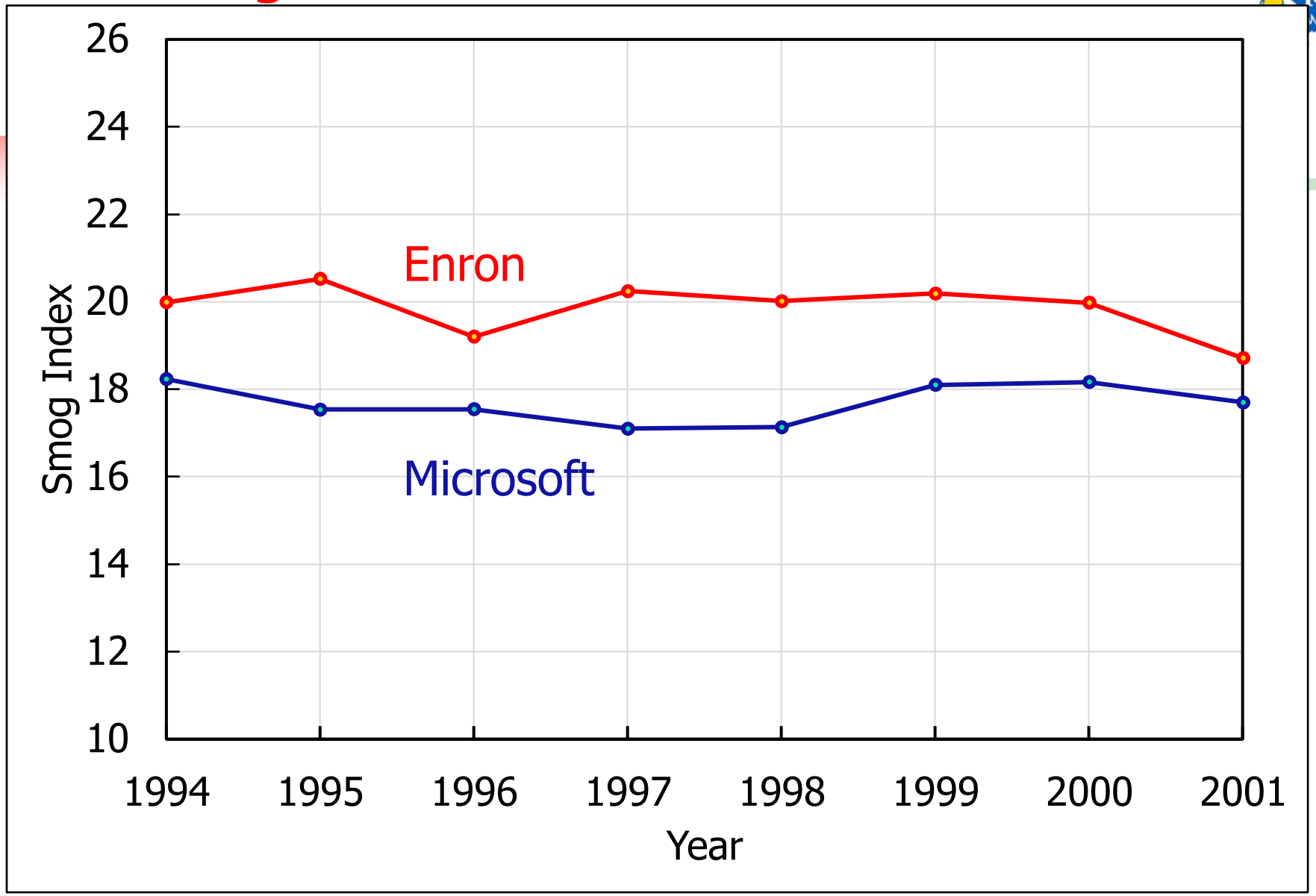
Flesch-Kincaid Readability Index for Satyam and WIPRO



Gunning-Fog Index for Enron and Microsoft for 10K (Annual Reports)



Smog Index for Enron and Microsoft



Textual Analysis Features in SeekiNF

1 - 20 of 434 < >

Download Results

Accenture plc
CIK :1467373
SIC :7389
File Type :10-K
File Date :10-29-2019
Filename :0001467373-19-000339.txt

NET 1 UEPS TECHNOLOGIES INC
CIK :1041514
SIC :6099
File Type :10-K
File Date :10-25-2019
Filename :0001062993-19-004031.txt

APPlife Digital Solutions Inc
CIK :1755101
SIC :7374
File Type :10-KAMEND
File Date :10-23-2019
Filename :0001445866-19-001259.txt

AB INTERNATIONAL GROUP CORP.
CIK :1605331
SIC :6794
File Type :10-K
File Date :10-22-2019
Filename :0001663577-19-000384.txt

DPW Holdings, Inc.
CIK :896493
SIC :3679
File Type :10-KAMEND
File Date :10-16-2019
Filename :0001214659-19-006507.txt

GROW CAPITAL, INC.
CIK :1448558
SIC :3669
File Type :10-K
File Date :10-15-2019
Filename :0001445866-19-001243.txt

Accenture plc CIK: 1467373 TYPE: 10-K DATE: 10-29-2019

View File | Download Raw File | Cosine Similarity/Word Variation | Raw Non Dollar Values in CSV | Raw Dollar Values in CSV

transformational change programs, either for one or more functions or business units, or across their entire organization. We provide industry-specific consulting services, as well as functional and technology consulting services. Our functional and technology consulting services include finance and enterprise performance; supply chain and operations; talent and organization; customers and channels; applications and architecture advisory; and technology advisory. We help our clients with the digital transformation of industries, enhancing our consulting services with digital, cloud, cybersecurity, artificial intelligence, blockchain and other capabilities.

Accenture Digital

innovate new services, capabilities and platforms through early adoption of technologies such as artificial intelligence, blockchain, machine learning, intelligent automation, extended reality and quantum computing to enhance productivity and create new growth opportunities.

- **Technology Innovation & Ecosystem. We harness**

future adversely affect client demand for our services and solutions. Our success depends, in part, on our ability to continue to develop and implement services and solutions that anticipate and respond to rapid and continuing changes in technology and offerings to serve the evolving needs of our clients. Examples of areas of significant change include digital-, cloud- and security-related offerings, which are continually evolving, as well as developments in areas such as artificial intelligence, augmented reality, automation, blockchain, Internet of Things, quantum computing and as-a-service solutions. Technological developments may materially affect the cost and use of technology by our clients and, in the case of as-a-service solutions, could affect the nature of how we generate revenue. Some of these technologies have reduced and replaced some of our historical services and solutions and may continue to do so in the future. This has caused, and may in the future cause, clients to

Word Distribution:

Download Distribution

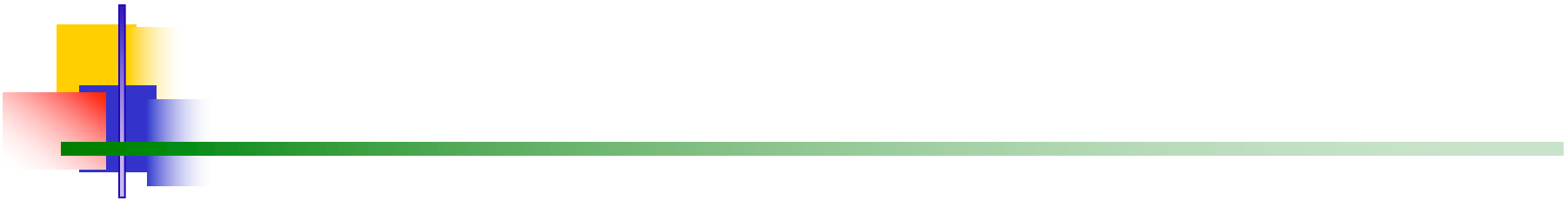
Download File Analysis

FILE ANALYSIS	
Total Word Count (TWC)	68965
Total Word Count without numerics (TWCWN)	64244
Sentence Count	1909
Gunning-Fog Index	24.4
Smog Index	20.801
Flesch Reading Ease	12.505
Flesch-Kincaid Grade Level	19.883
Automated Readability Index	21.132
Coleman-Liau Index	16.364

RISK SENTIMENT ANALYSIS			
RISK TYPE	COUNT (NR)	LN(1+NR)	NR PERCENT TWC
RS Feng Li	109	4.7	0.158
RS Financial	243	5.497	0.352
RS Litigation	236	5.468	0.342
RS Tax	332	5.808	0.481
RS Systematic	553	6.317	0.802
RS Idiosyncratic	1009	6.918	1.463
RS Overall	2373	7.772	3.441

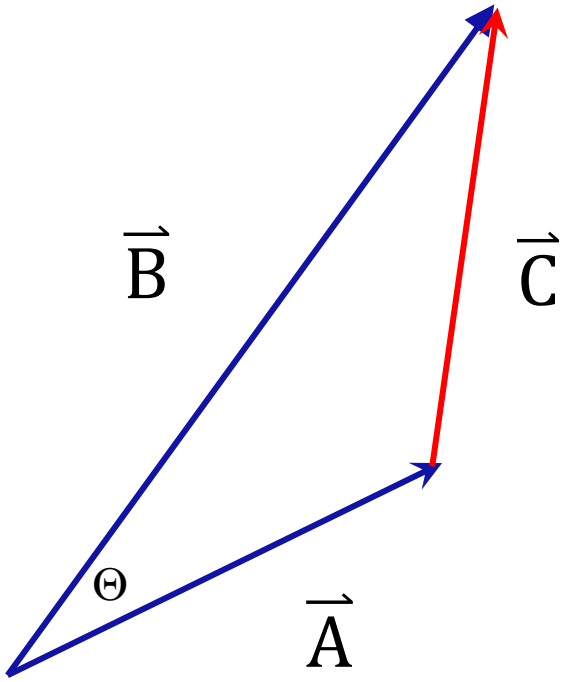
SeekEdgar

Analytical Geometry Applied to Textual Analysis



Cosine Measure of Similarity And Vector Variation Score

Cosine Measure of Similarity & Vector Variation Score



$$\vec{C} = \vec{B} - \vec{A}$$

$$\text{Cos}(\Theta) = \vec{A}.\vec{B}/(|A|.|B|)$$

Vector Variation Score = Normalized $|\vec{C}|$
$= |\vec{B} - \vec{A}|/(|A| + |B|)$

Cosine Measure of Similarity Between $\vec{A}$ and $\vec{B}$ = Cos(Θ);
Θ = Angle between $\vec{A}$ and $\vec{B}$

Cosine Measure of Similarity Based on Textual Analysis

- Each Word/Phrase is considered to be an axis
  - Words including numbers
  - Words without numbers
  - Words without numbers and stop words
  - Specific Phrases or words
- Problems with Cosine Measure of Similarity
  - Cannot distinguish between "mike is eating apple" and "apple is eating mike". They are identical under CMS.
  - Use of Natural Language Processor (NPL) to distinguish such a situation
  - Cannot distinguish between the two documents, if one document is just the multiple of each word in another document. Use of Vector Variation Score helps in such a situation.

Some Applications of Cosine Measure of Similarity

- Determining how similar two documents are
- How to classify peer groups and competitors based on descriptions of product attributes and strategies
- Find candidates for mergers and acquisitions based on descriptions of product attributes and strategies
- Assess various business risks by comparing the occurrences of various risk factors

SeekiNF

Anytime, Anywhere...

SEC filings at your fingertips in seconds with See

HOME | ABOUT US | SeekiNF | FRAANK | FAQs | CONTACT US | Welcome Srivastava

Recent Press and Other Releases | 10-K Exhibit 21(Subsidiaries) | Conference Call Transcripts | Search | Request Form | Special Request | Guidelines and Examples

CIK : 1275158 Change CIK

Add CIK

Year : 1994 to 2019

File Type:

10-K	8-K	20-F	Form-3	424B1	424B4	15F-12B	DEF 14A	N-CSR	UPLOAD	MDNA 10-K
10-Q	6-K	40-F	Form-4	424B2	424B5	15F-12G	DEFM14A	NSAR-A	CORRESP	MDNA 10-Q
N-Q	S-1	11-K	Form-5	424B3	424B7	15F-12D	DEFM14C	NSAR-B	ITEM 1A	FOOTNOTE 10-K
SD	S-4	15-12B	15-12G	15-12D	424B8	SC 13D	SC 13G	NSAR-U	13F-HR	FOOTNOTE 10-Q

Cosine Similarity measures how close two documents are. **Word Variation** compares and provides the frequency of words that appear in two documents.

Search

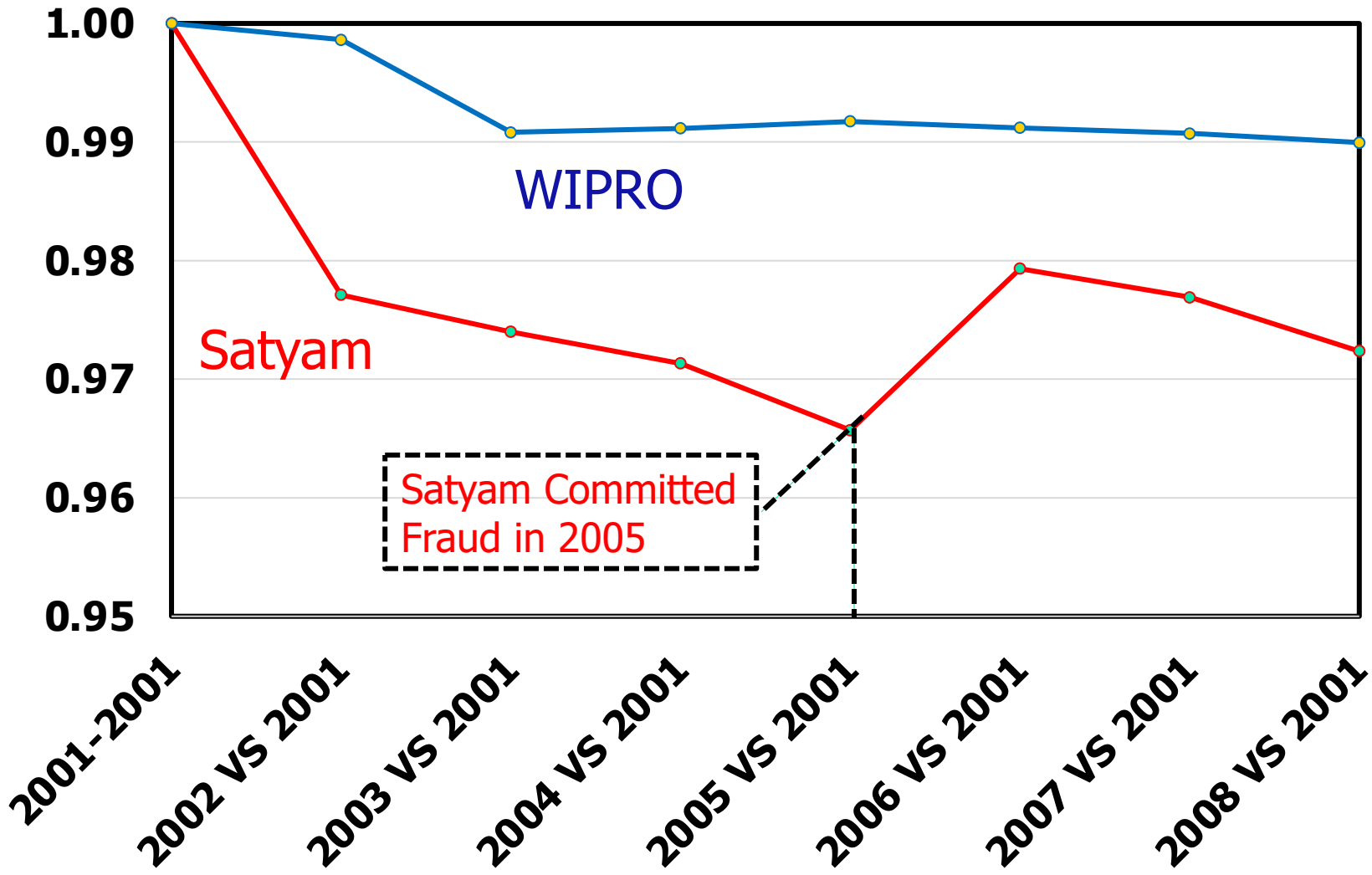
Trail Version: Please email Tech Team at techteam@seekedgar.com if any error or suggestions.

The list of 10-K 's filed between 1994 - 2019 years for CIK : 1275158

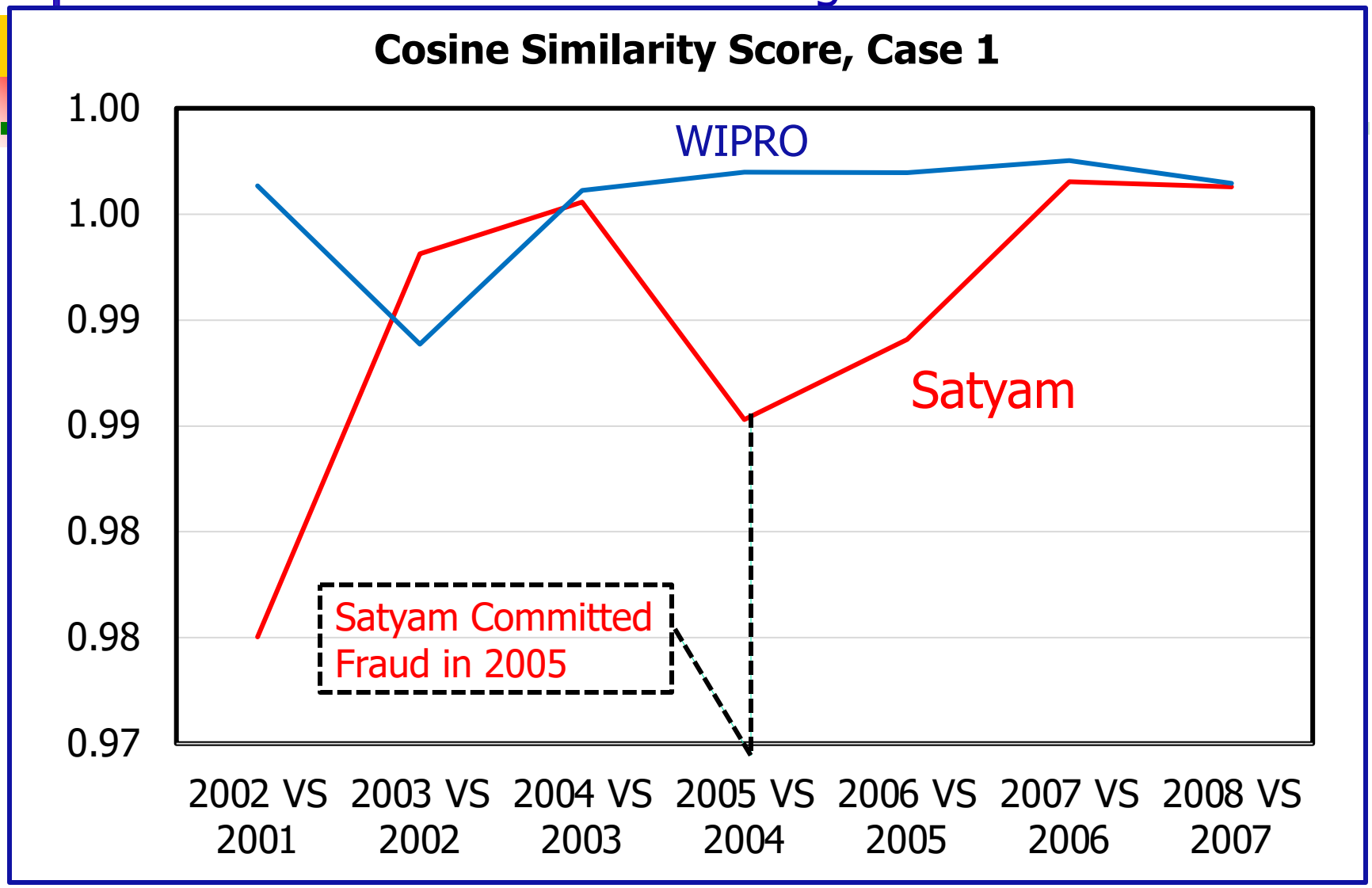
Cosine Measure of Similarity

Cosine Similarity			Word Variation	
Filing	**SEC File Link**	**Filing Date**	**Word Distribution**	
10-K	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-19-000010-index.htm	03-15-2019	Download Distribution	
10-K	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-18-000017-index.htm	03-15-2018	Download Distribution	
10-K	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-17-000018-index.htm	03-02-2017	Download Distribution	
10-K	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-16-000096-index.htm	03-01-2016	Download Distribution	
10-K/A	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-15-000026-index.htm	05-07-2015	Download Distribution	
10-K	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-15-000005-index.htm	02-24-2015	Download Distribution	
10-K	https://www.sec.gov/Archives/edgar/data/1275158/0001275158-14-000011-index.htm	03-07-2014	Download Distribution	

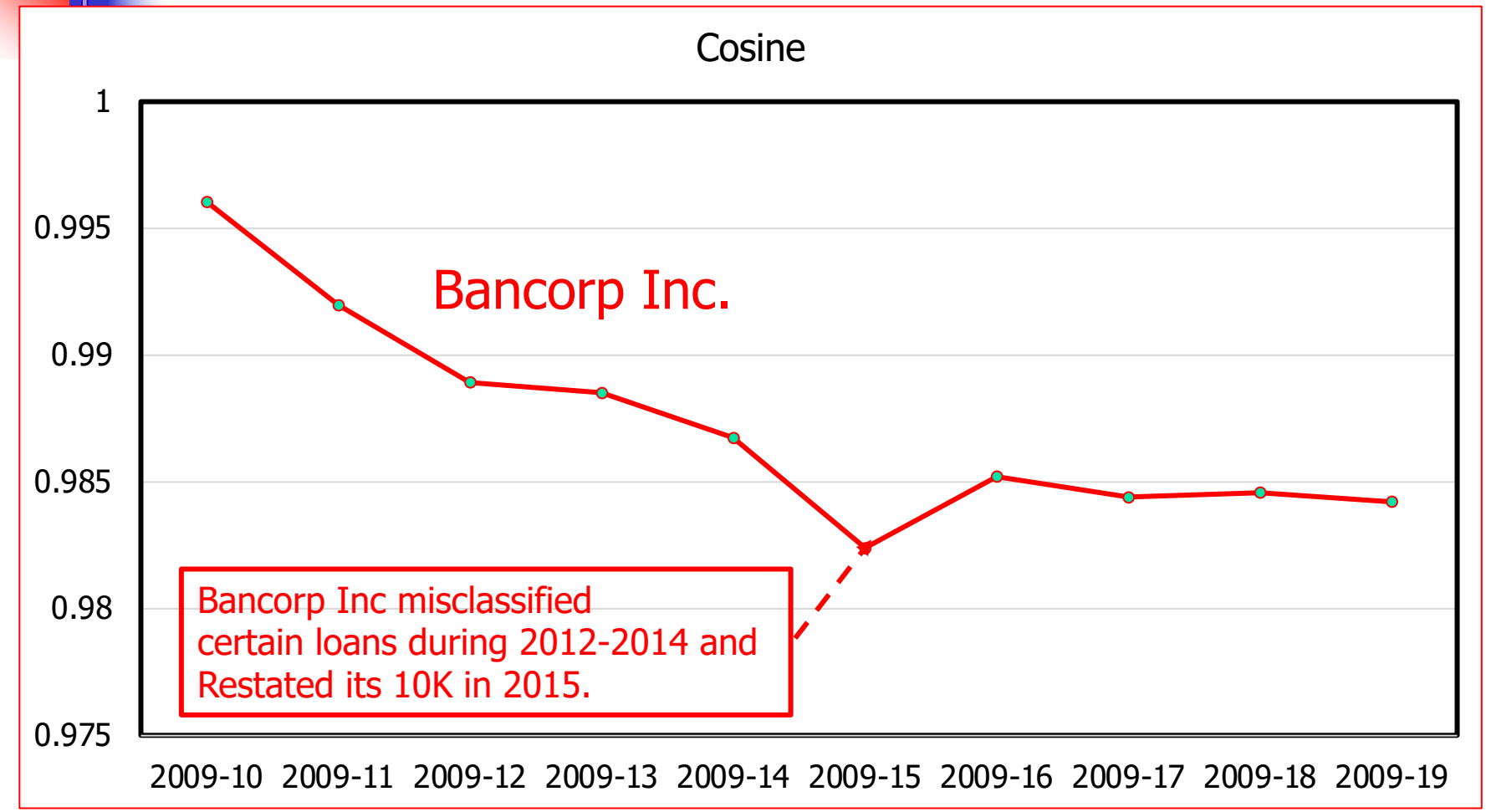
Cosine Similarity Score for Satyam and WIPRO in Reference to 2001 20F with all the Words Including Numbers



Subsequent Changes in Cosine Similarity for Satyam and WIPRO for 20-F with all the words including Numbers



Graph of Cosine Measure of Similarity for Bancorp Inc. with respect to 2009 10K

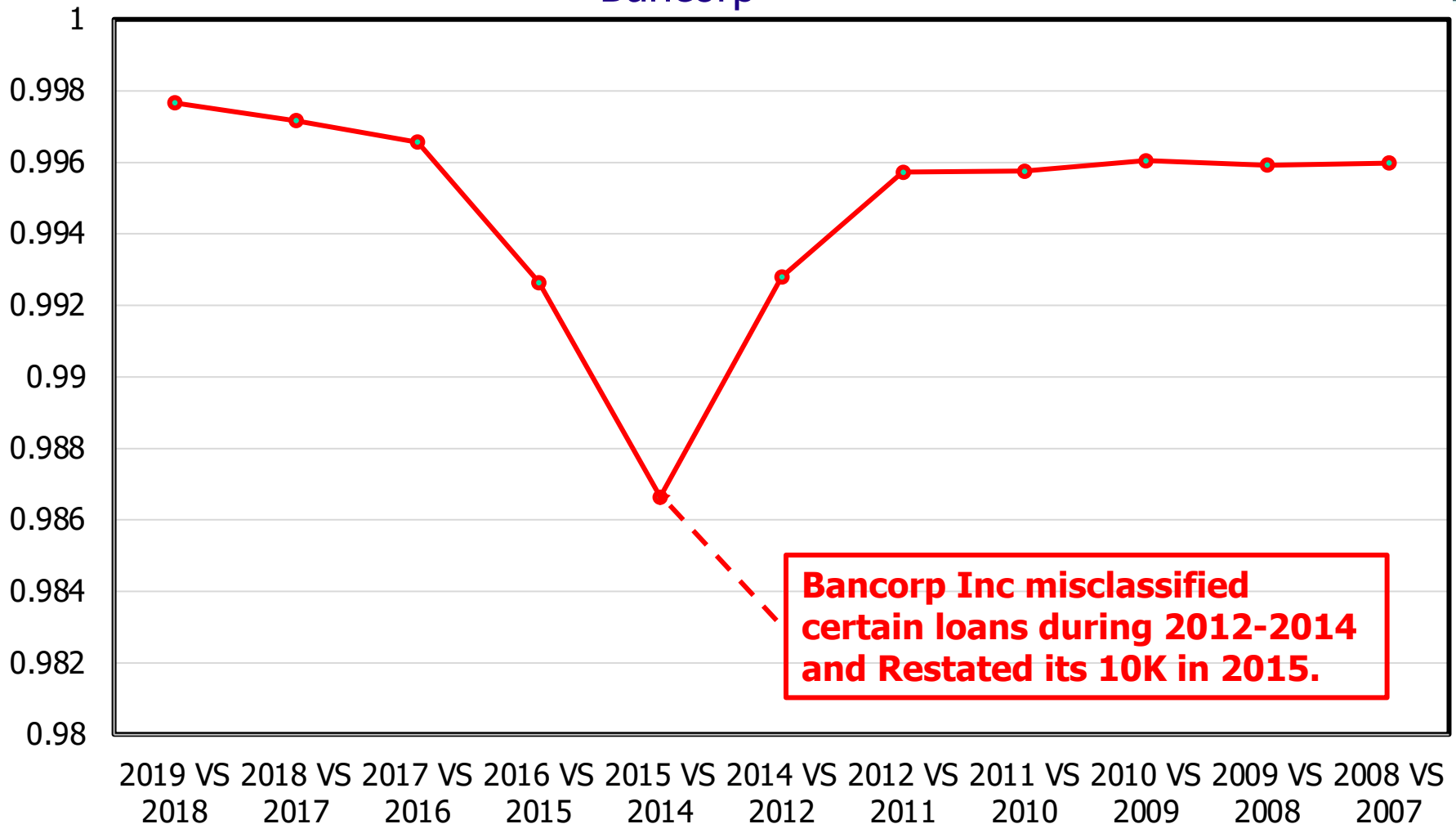


Cosine Similarity Score For:	Bancorp, Inc.										
FILE TYPE :	10K										
	File 1 VS File 2	File 2 VS File 3	File 3 VS File 4	File 4 VS File 5	File 5 VS File 6	File 6 VS File 7	File 7 VS File 8	File 8 VS File 9	File 9 VS File 10	File 10 VS File 11	File 11 VS File 12
CIK	2019 VS 2018	2018 VS 2017	2017 VS 2016	2016 VS 2015	2015 VS 2014	2014 VS 2012	2012 VS 2011	2011 VS 2010	2010 VS 2009	2009 VS 2008	2008 VS 2007
1295401	0.997664077	0.997168298	0.99657023	0.99263956	0.9866455	0.99279141	0.99572869	0.99575684	0.99604947	0.99592561	0.9959804

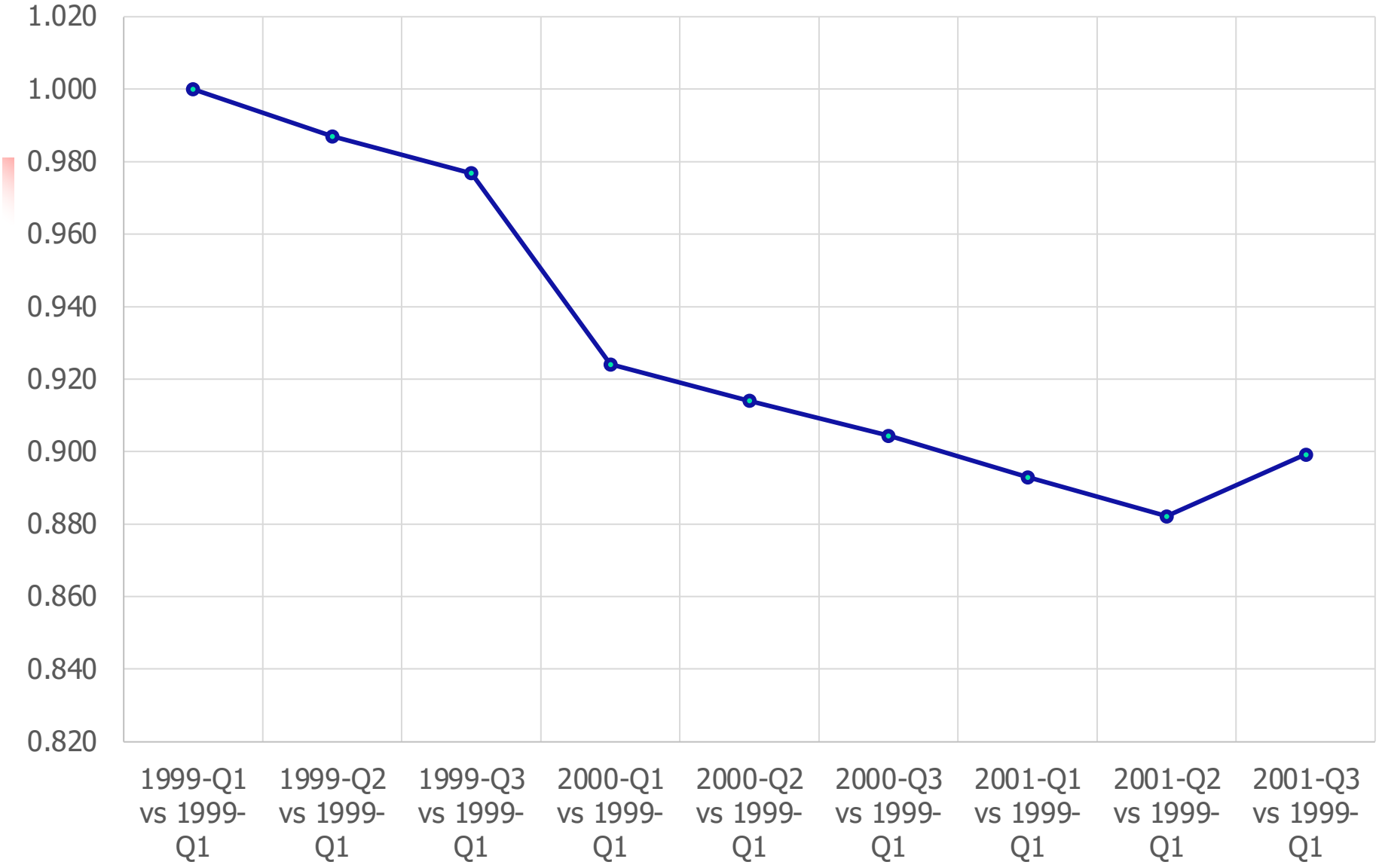
File 1	https://www.sec.gov/Archives/edgar/data/1295401/0001295401-19-000040-index.html
File 2	https://www.sec.gov/Archives/edgar/data/1295401/0001562762-18-000100-index.html
File 3	https://www.sec.gov/Archives/edgar/data/1295401/0001295401-17-000002-index.html
File 4	https://www.sec.gov/Archives/edgar/data/1295401/0001295401-16-000017-index.html
File 5	https://www.sec.gov/Archives/edgar/data/1295401/0001295401-15-000005-index.html
File 6	https://www.sec.gov/Archives/edgar/data/1295401/0001295401-14-000003-index.html
File 7	https://www.sec.gov/Archives/edgar/data/1295401/0000950159-12-000151-index.html
File 8	https://www.sec.gov/Archives/edgar/data/1295401/0000950159-11-000137-index.html
File 9	https://www.sec.gov/Archives/edgar/data/1295401/0000950159-10-000205-index.html
File 10	https://www.sec.gov/Archives/edgar/data/1295401/0001193125-09-059928-index.html
File 11	https://www.sec.gov/Archives/edgar/data/1295401/0001193125-08-059003-index.html
File 12	https://www.sec.gov/Archives/edgar/data/1295401/0001193125-07-056963-index.html

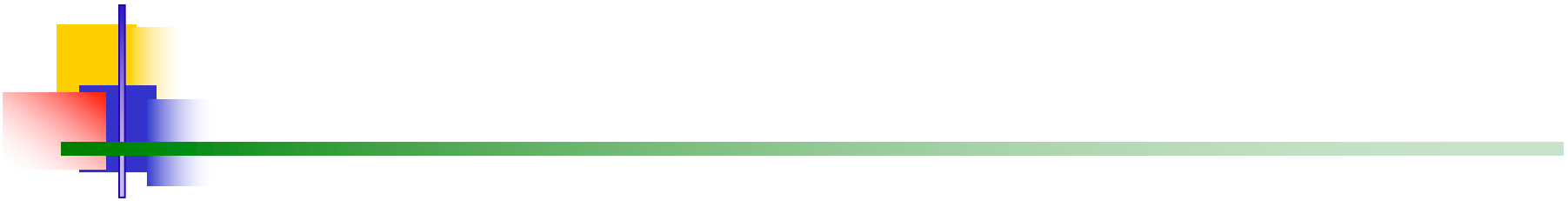
Cosine Years	Cosine
2019 VS 2018	0.997664077
2018 VS 2017	0.997168298
2017 VS 2016	0.996570225
2016 VS 2015	0.992639558
2015 VS 2014	0.986645505
2014 VS 2012	0.992791407
2012 VS 2011	0.995728688
2011 VS 2010	0.99575684
2010 VS 2009	0.99604947
2009 VS 2008	0.995925613
2008 VS 2007	0.995980397

Subsequent Change in Cosine Similarity Measure for Bancorp



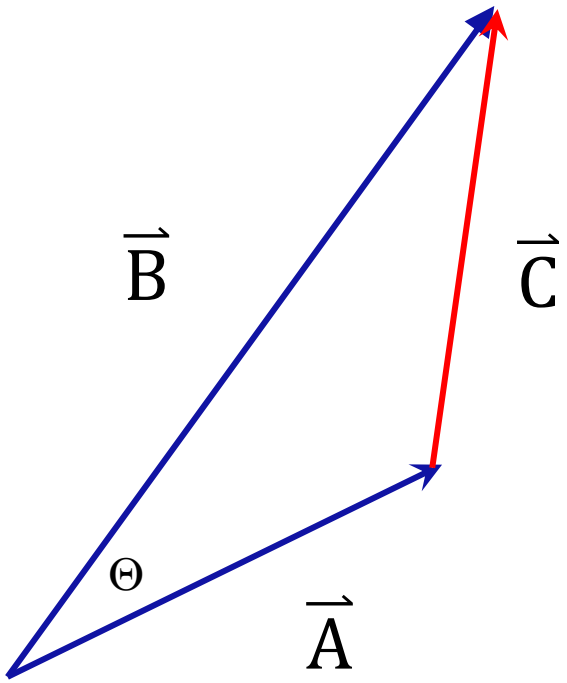
Enron -Cosine Measure with respect to 1999 10Q1





Cosine Measure of Similarity Versus Vector Variation Score

Cosine Measure of Similarity & Vector Variation Score



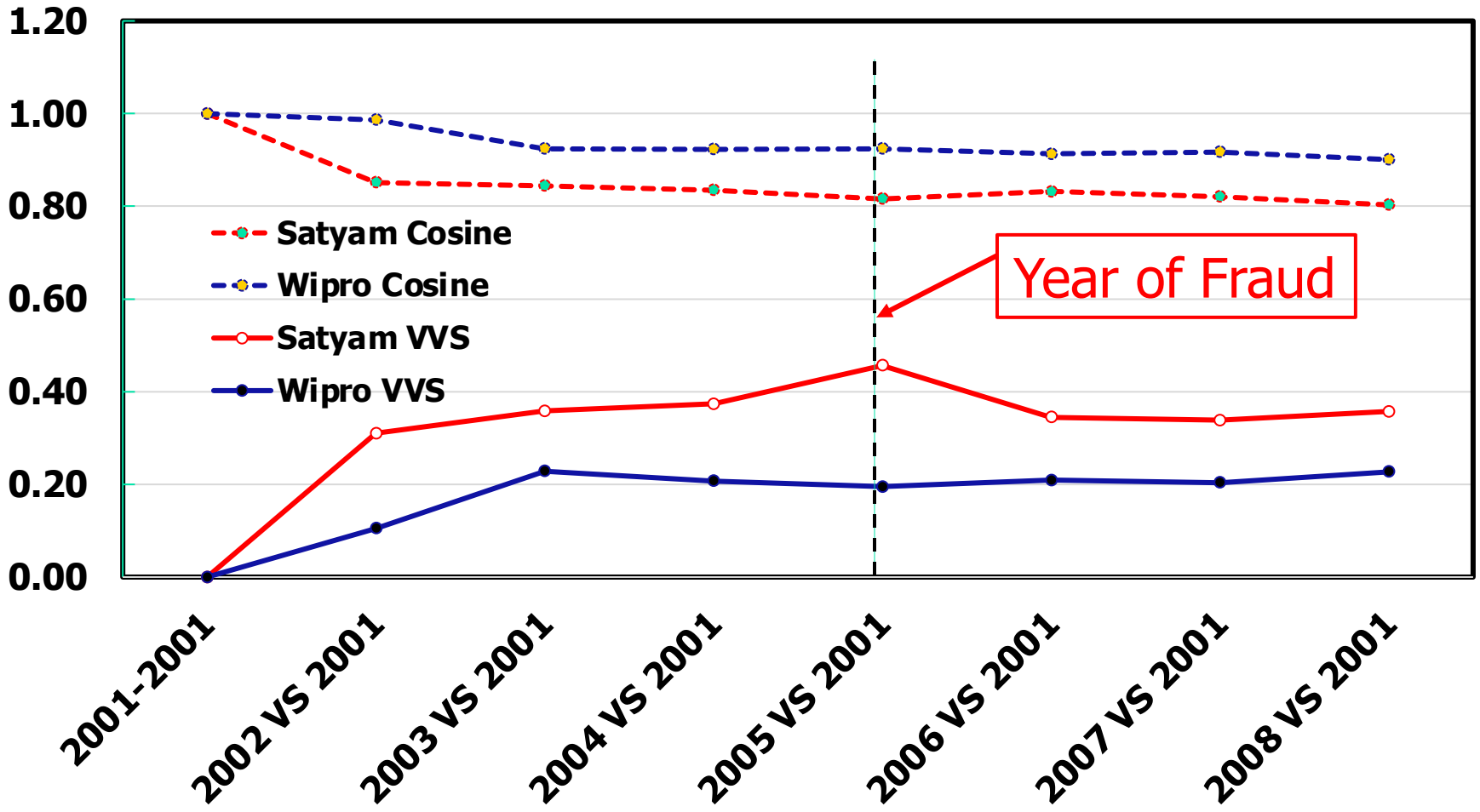
$$\vec{C} = \vec{B} - \vec{A}$$

$$\text{Cos}(\Theta) = \vec{A}.\vec{B}/(|A|.|B|)$$

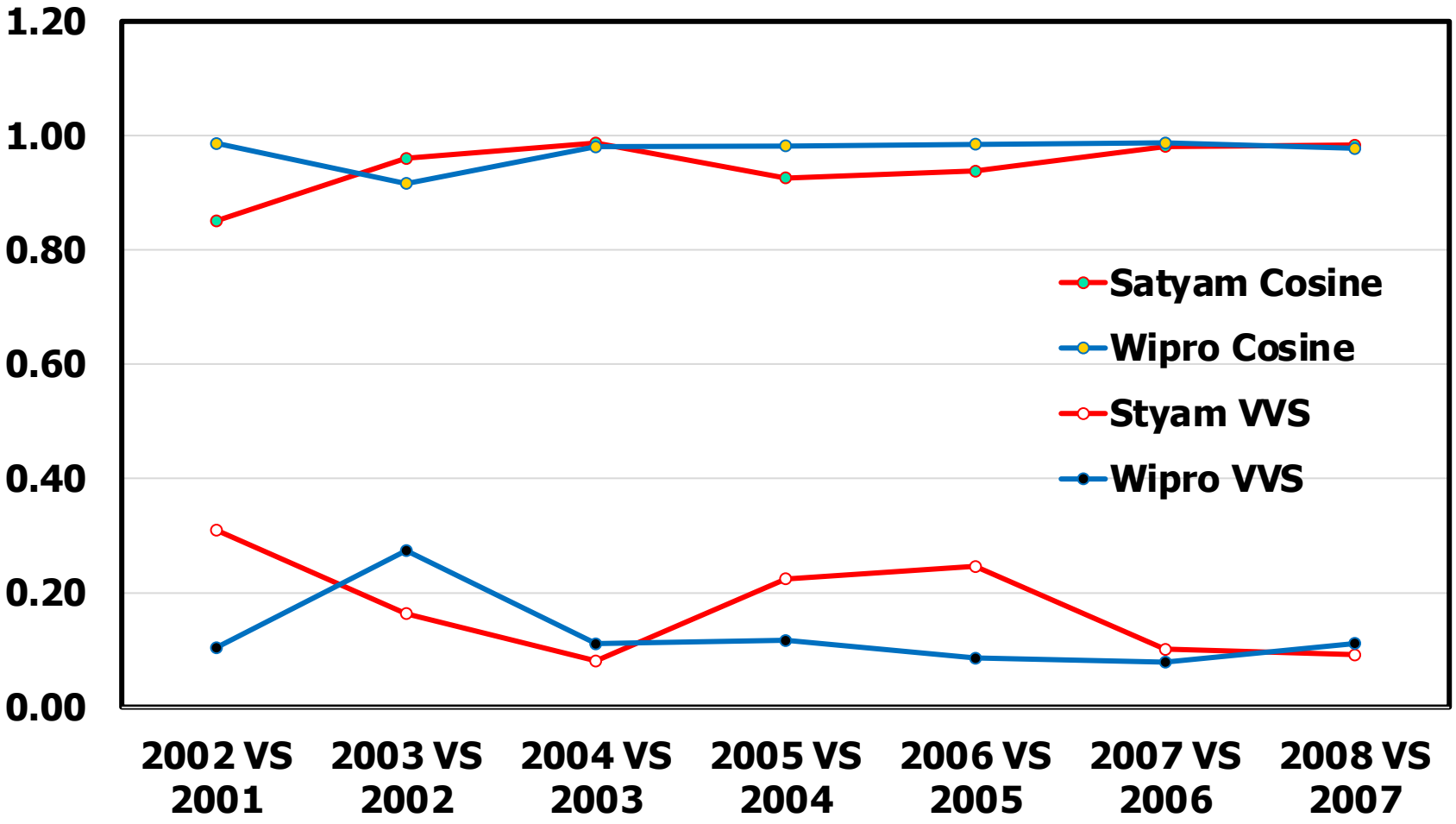
Vector Variation Score = Normalized $|\vec{C}|$
$= |\vec{B} - \vec{A}|/(|A| + |B|)$

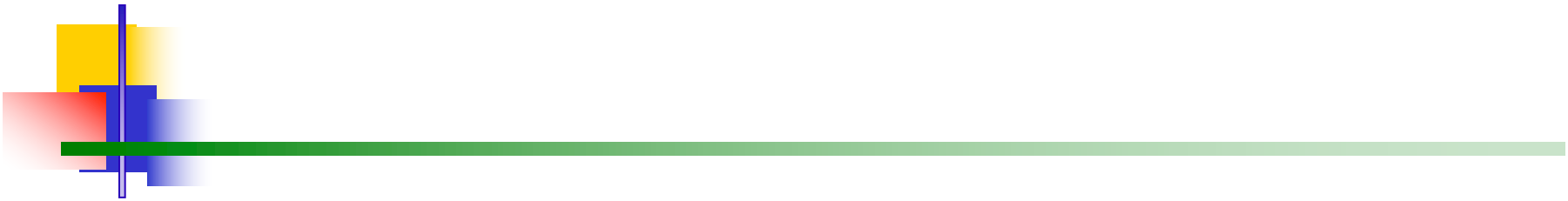
Cosine Measure of Similarity Between $\vec{A}$ and $\vec{B}$ = Cos(Θ);
Θ = Angle between $\vec{A}$ and $\vec{B}$

Cosine Similarity and Vector Variation Scores for Satyam and Wipro for 20F in Reference to 2001 with Words without Numbers and Stop Words.



Subsequent Changes in Cosine Similarity and Vector Variation Scores for Satyam and Wipro for 20F with Words without Numbers and Stop Words.





Word Variation From One Year to Another

Enron

	https://www	https://www	https://www	https://www	https://www	https://www	https://www	https://www
WORDS	1024401 -- 2001	1024401 -- 2000	1024401 -- 1999	1024401 -- 1998	72859 -- 1997	72859 -- 1996	72859 -- 1995	72859 -- 1994
its	52	50	77	43	21	24	29	15
wholesale	52	54	35	24	0	0		0
on	51	41	79	41	14			16
market	48	42	34	27				10
related	47	29	39				14	
increased	45	37				47		
other	45				28			
value				22				2
			39				7	
		44	33			11	16	
		38				48		
	36				11			15
equity				24			14	7
			22			28	43	47
		32				29	42	32
	33	26			29	23	18	20
management	32			27	16	26	24	13
primarily			35	43	40	45	51	51
		36	19	17	9	8	3	5
		26	25	17	20	20	32	25
	27	9	0	0	0	0	0	0
markets	26	22	14	14	15	10	4	1
business	24	13	48	19	8	10	15	16

Enron Wholesale operates in developed markets such as North America and Europe , as well as developing

Broadband Services also makes investments in companies with related technologies and with the potential for capital appreciation.

Enron Wholesale operates in developed markets such as North America and Europe , as well as developing or newly deregulating markets including South America , India

Average Reduced Frequency (ARF)

i.e.,

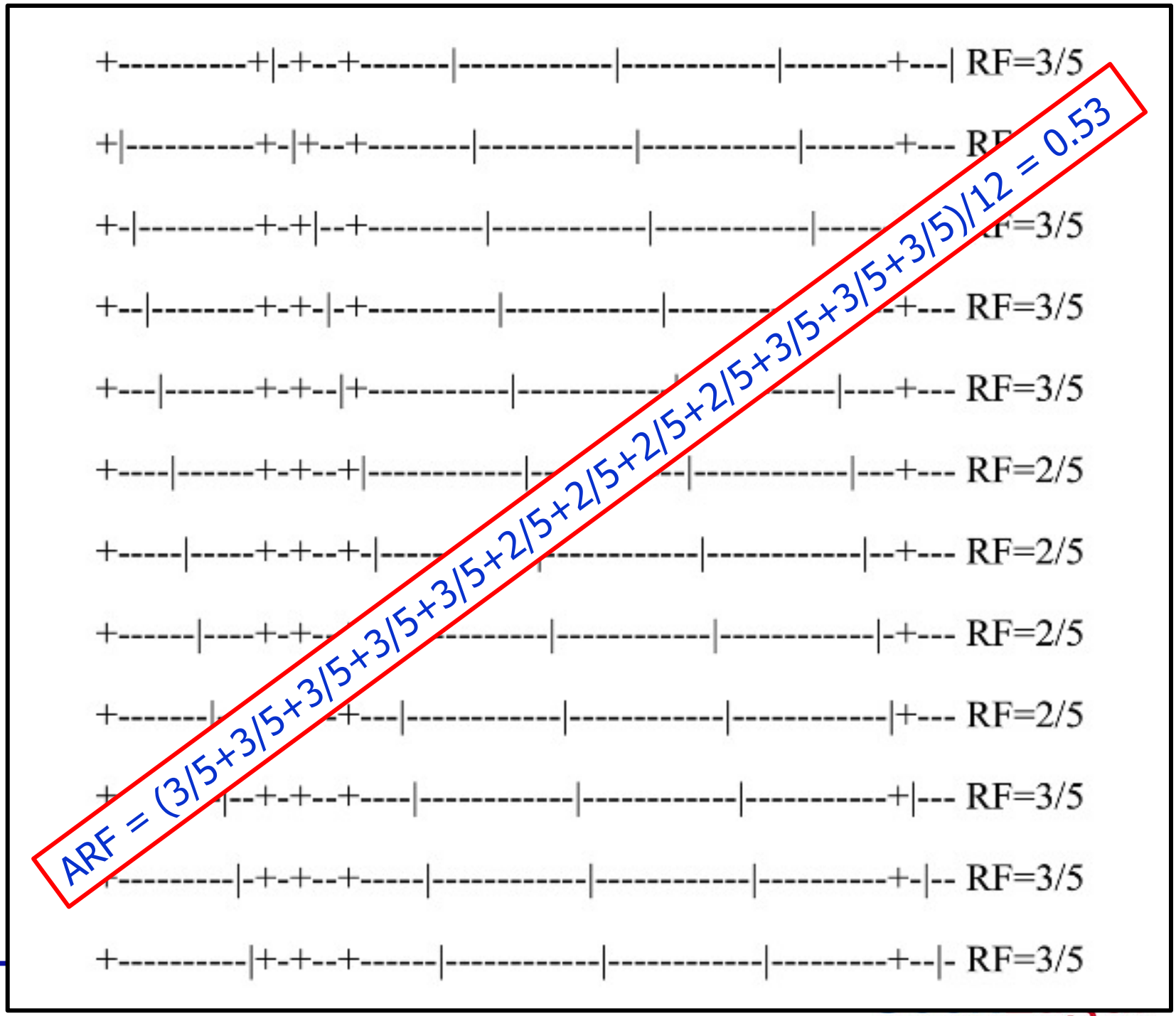
Tone Dispersion

Illustration of Average Reduced Frequency (ARF) Calculation

- Assume a document containing 60 words, 5 of which are positive tone words. Let us represent positive tone words by the plus sign (+),and non-tone words by a hyphen (-).
- The entire document is divided into five corpus each containing 12 words.
- Five Positive tone words are at locations: 0, 11, 13, 16, 56.
- +----------+|-+--+-------|-----------|-----------|---------+---| RF=3/5
- RF represents the number of corpus containing the Tone words divided by the total number of corpus.
- Calculate various RFs for the cases where each subsequent corpus distribution is moved one word towards right, and so on.

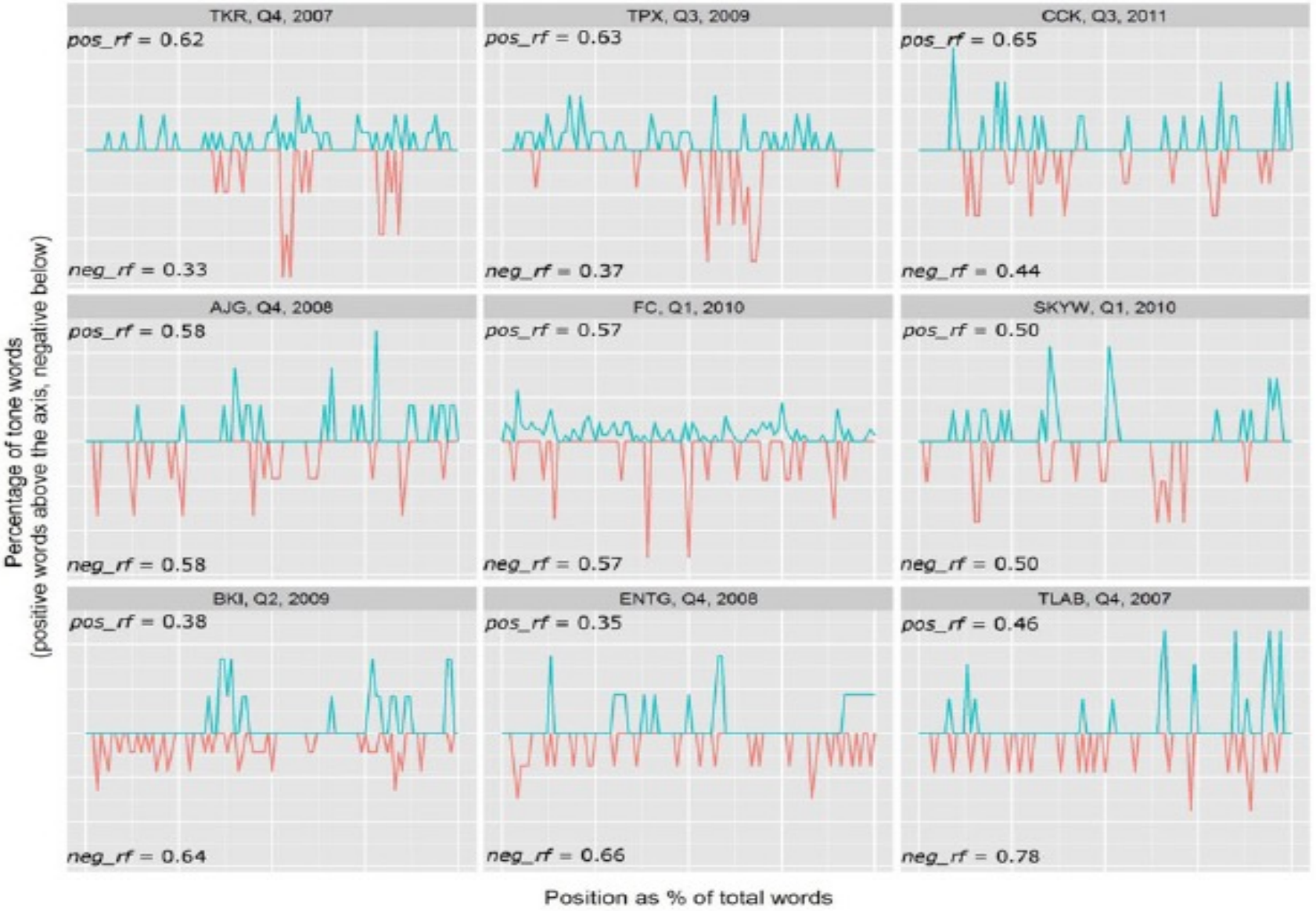
ADJUSTED TONE DISPERSION (Average Reduced Frequency, ARF)

Allee and De Angelis, JAR (2015)



The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion

(K. D. ALLEE and M. D . DEANGELIS, JAR Vol. 53, No. 2, 2015, p.241)



Risk Sentiment (RS) of 10K by Feng Li

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=898181

"Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?"

Important Findings

1. Increase in RS is associated with Lower Future Earnings
2. Firms with a larger increase in risk sentiment have more negative earnings changes in the next year
3. Firms with a large increase in RS experience significantly negative returns relative to those firms with little increase in risk sentiment in the twelve months after the annual report filing date.

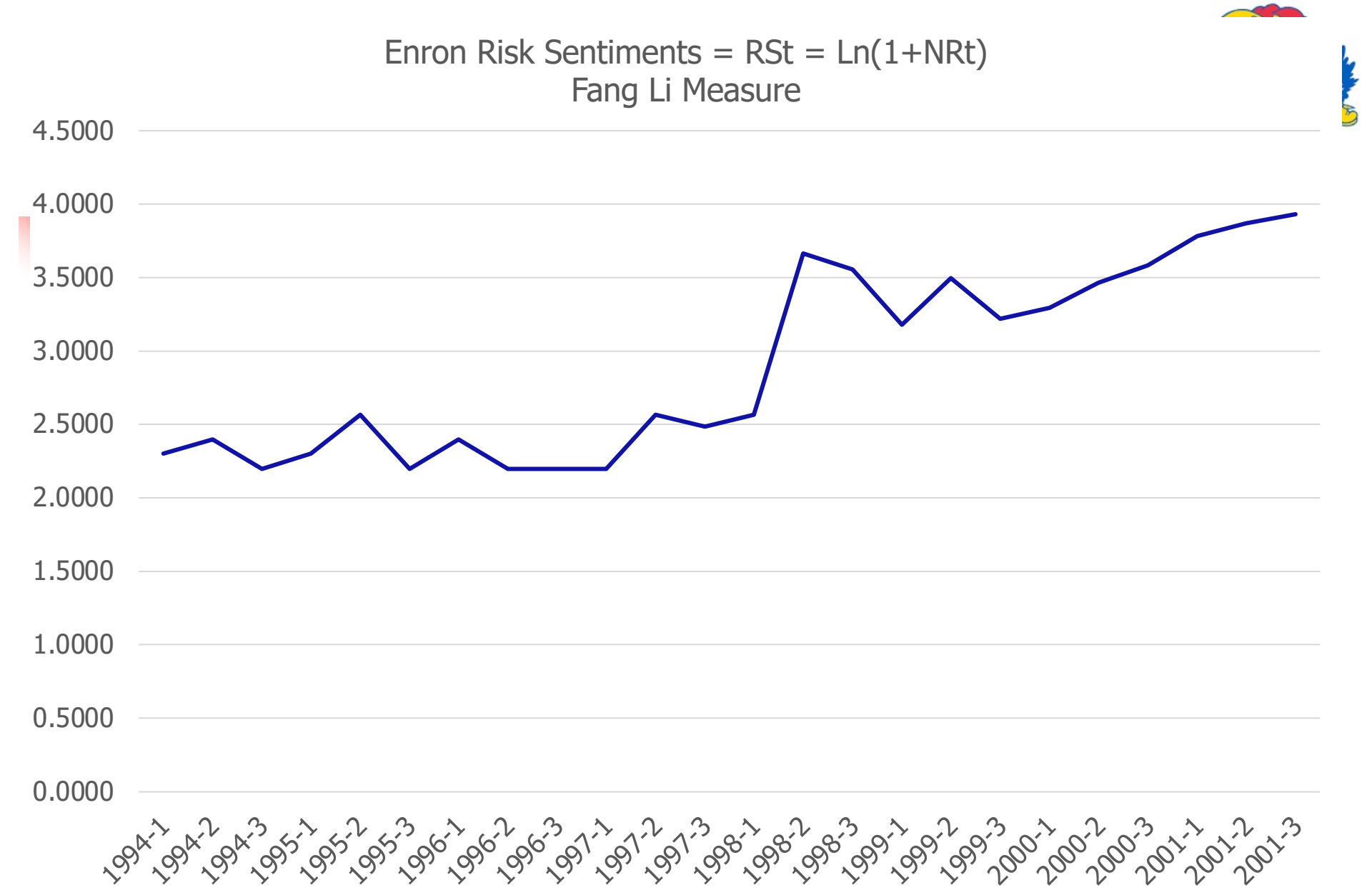
Risk Sentiment measure by Feng Li

- Definition of Risk Sentiment:
  - $RS_t = \ln(1+NR_t)$

where NR_t is the numbers of occurrence of risk-related words in year t.

- The key words used to measure the Risk Sentiment of 10K are:
  - risk", "risks", "risky", "uncertain", "uncertainty", and "uncertainties
- Change of risk sentiment as
  - $\Delta RS_t = \ln(1+NR_t) - \ln(1 + NR_{t-1})$

where NR_t and NR_{t-1} are the numbers of occurrence of risk-related words in year t and year t – 1 respectively.


Enron Risk Sentiments = RSt = Ln(1+NRt)
Fang Li Measure
4.5000
4.0000
3.5000
3.0000
2.5000
2.0000
1.5000
1.0000
0.5000
0.0000
1994-1
1994-2
1994-3
1995-1
1995-2
1995-3
1996-1
1996-2
1996-3
1997-1
1997-2
1997-3
1998-1
1998-2
1998-3
1999-1
1999-2
1999-3
2000-1
2000-2
2000-3
2001-1
2001-2
2001-3

Risk Sentiments using Item 1A

by Campbell, Chen, Dhaliwal, Lu, and Steele

https://link.springer.com/article/10.1007/s11142-013-9258-3

- Financial Risk
- Litigation Risk
- Tax Risk
- Idiosyncratic Risk (specific to company)
- Systematic Risk (economy wide risk)
- Overall Risk

The information content of mandatory risk factor disclosures in corporate filings

(Item 1A)

by

John L. Campbell • Hsinchun Chen •
Dan S. Dhaliwal • Hsin-min Lu • Logan B. Steele

In

Rev Account Stud (2014) 19:396–455

Word List for Financial Risk

Table 9 Key words list by risk category

Risk category	Keyword	Risk category	Keyword
Financial	Anti-takeover (provisions|provision)	Financial	Reserves
Financial	Bank debt	Financial	Revolver
Financial	Capital (expenditure|expenditures)	Financial	Sale of productive assets
Financial	Capital (lease|leases)	Financial	Stock market listing
Financial	Chapter 11	Financial	Stock price drop
Financial	Chapter 7	Financial	Stock price volatility
Financial	Chapter 9	Financial	Underfunded pensions
Financial	Collateral	Financial	Underwriting
Financial	Concentrated ownership	Financial	Volatility of operating results
Financial	(Covenant|covenants)	Financial	Volatility of revenues
Financial	Credit (facility|facilities)	Financial	Volatility of sales
Financial	Credit rating	Financial	Working capital
Financial	Credit risk	Other-Idiosyncratic	Acquisition
Financial	Debt burden	Other-Idiosyncratic	Adequate staffing
Financial	Decline in stock price	Other-Idiosyncratic	Advertising

Word List for Litigation Risk

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Legal and Regulatory	Pending (lawsuit|lawsuits)	Other-Systematic	Foreign exchange
Legal and Regulatory	Plaintiff	Other-Systematic	(Forward|forwards)
Legal and Regulatory	Possibility of (restatement|restatements)	Other-Systematic	Fuel
Legal and Regulatory	Potential (lawsuit|lawsuits)	Other-Systematic	Future
Legal and Regulatory	Product liability	Other-Systematic	Gas
Legal and Regulatory	(Regulation|regulations)	Other-Systematic	Gasoline
Legal and Regulatory	Regulatory	Other-Systematic	GDP
Legal and Regulatory	Regulatory approval	Other-Systematic	G.D.P.
Legal and Regulatory	Regulatory change	Other-Systematic	GNP
Legal and Regulatory	Regulatory compliance	Other-Systematic	G.N.P.
Legal and Regulatory	Regulatory environment	Other-Systematic	General business risks
Legal and Regulatory	Related (party|parties)	Other-Systematic	General conditions

Word List for Litigation Risk

Tax	Aggressive tax (position|positions)
Tax	Back taxes
Tax	Deferred tax (asset|assets)
Tax	Deferred tax (liability|liabilities)
Tax	Excise (tax|taxes)
Tax	FIN 48
Tax	Internal Revenue Service
Tax	IRS
Tax	I.R.S.
Tax	IRS audit
Tax	IRS judgment
Tax	Loss (carryback|carrybacks)
Tax	Loss (carryforward|carryforwards)
Tax	Property (tax|taxes)
Tax	Provision for income (tax|taxes)
Tax	State (tax|taxes)
Tax	(Tax|Taxes)
Tax	Tax audit
Tax	Tax (authority|authorities)
Tax	Tax (liability|liabilities)
Tax	Tax (penalty|penalties)
Tax	Taxable

Risk Sentiment Metrics using SeekiNF with Built-in Features

- Risk Sentiment (Feng Li Model)
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=898181
- Risk Sentiments (Campbell et al. Model)
  https://link.springer.com/article/10.1007/s11142-013-9258-3
  - Risk Sentiment (Financial)
  - Risk Sentiment (Legal and Regulatory, i.e., Litigation)
  - Risk Sentiment (Tax)
  - Risk Sentiment (Systematic, economy)
  - Risk Sentiment (Idiosyncratic, specific to firm)
  - Risk Sentiment (Overall)

Risk Sentiment Metrics using SeekiNF with Built-in Features

- Step 1: Type "a|the|of" without quotes
- Step 2: Input a set of companies CIKs or leave the default "All" and select the period
- Step 3: Select the desired resolution from "Paragraph (All)", "Footnote" and "MD&A"
- Step 4: Select the Filings desired from, 10K, 10Q, 20F and 40F and their amendments.

SeekEdgar

Anytime, Anywhere...

SEC filings at your fingertips in seconds with SeekiNF

STEP 1: Please enter Phrase(s)/Keyword(s) you wish to search:

With the exact phrase(s): a|the|of

The following features are for Single Phrase/Word only: (In case of Multiple Phrases/Words, please use the Request Form to submit the request.)

- [ ] Phrase/Word Counter
- [ ] Display words before & after

Number of words before: Number of words After:

Proximity Search:

Please input within how many words:

With all of the words:

With at least one of the words:

Without the words:

STEP 2: Please select the options below (Default - ALL):

COMPANY: All, Name

From: 1994 To: 2020

SIC (Optional): All SICs

STEP 3: Please select the search Resolution here:

- (•) Paragraph(All)
- ( ) Footnote
- ( ) SOX 404 Mgt Report
- ( ) MD&A
- ( ) Table
- ( ) Audit Report
- ( ) SOX 404 Audit Report

STEP 4: Please select the Document(s) you want to search here:

( ) All

(•) Select (Please see Filing Description for more information): ( ) Un-Check All

Annual Reports (•) Check All Annual Reports ( ) Un-Check All Annual Reports

- [x] 10-K [x] 10-KT [x] 10-K405 [x] 10KT405 [x] 10KSB [x] 10KSB40
- [x] 10-K/A [x] 10-KT/A [x] 10-K405/A [x] 10KT405/A [x] 10KSB/A [x] 10KSB40/A

Quarterly, Current & Foreign Annual Reports

- [ ] 10-Q [ ] 10-QT [ ] 10QSB [ ] 8-K [ ] 6-K [ ] 20-F [ ] 40-F
- [ ] 10-Q/A [ ] 10-QT/A [ ] 10QSB/A [ ] 8-K/A [ ] 6-K/A [ ] 20-F/A [ ] 40-F/A

Proxies & Registrations

- [ ] DEF 14A [ ] DEFM14A [ ] S-1 [ ] S-4 [ ] 15-12B [ ] 15-12G
- [ ] DEFA14A [ ] DEFM14C [ ] S-1/A [ ] S-4/A [ ] 15-12B/A [ ] 15-12G/A
- [ ] 15F-12B [ ] 15F-12G [ ] ABS-15G [ ] PRE 14A
- [ ] 15F-12B/A [ ] 15F-12G/A [ ] ABS-15G/A

Ownership & Prospectuses

- [ ] Form 3 [ ] Form 4 [ ] Form 5 [ ] 424B1 [ ] 424B3 [ ] 424B5
- [ ] Form 3/A [ ] Form 4/A [ ] Form 5/A [ ] 424B2 [ ] 424B4 [ ] 424B7
- [ ] 13F-HR [ ] 13F-NT [ ] SC 13D [ ] SC 13G [ ] 424B8
- [ ] 13F-HR/A [ ] 13F-NT/A [ ] SC 13D/A [ ] SC 13G/A

Other Filings

- [ ] N-CSR [ ] N-CSRS [ ] NSAR-A [ ] NSAR-AT [ ] NSAR-B
- [ ] N-CSR/A [ ] N-CSRS/A [ ] NSAR-A/A [ ] NSAR-AT/A [ ] NSAR-B/A
- [ ] NSAR-BT [ ] NSAR-U [ ] N-Q [ ] SD [ ] CORRESP (Comment Letters)
- [ ] NSAR-BT/A [ ] NSAR-U/A [ ] N-Q/A [ ] SD/A [ ] UPLOAD (Response Letters)
- [ ] AAER** [ ] 11-K [ ] 11-K/A [ ] CT ORDER

10-K, 10-Q & 8-K Exhibits & Shareholders Letters

- [ ] PRESS RELEASE* [ ] SHAREHOLDERS MEETINGS* [ ] CONFERENCE CALLS*
- [ ] OTHER 8-K EX [ ] SHAREHOLDERS LETTER* [ ] EXHIBIT 21 (10-K & 10-Q)
- [ ] EXHIBIT 95 (10-K & 10-Q) [ ] EXHIBIT 10 (10-K) [ ] EXHIBIT 10 (10-Q)
- [ ] EXHIBIT 99 (10-K) [ ] EXHIBIT 99 (10-Q) [ ] ITEM 1A (10-K)

Public Company Accounting Oversight Board

- [ ] INSPECTION REPORTS** [ ] SETTLED DISCIPLINARY ORDERS**
- [ ] ADJUDICATED DISCIPLINARY ACTIONS**

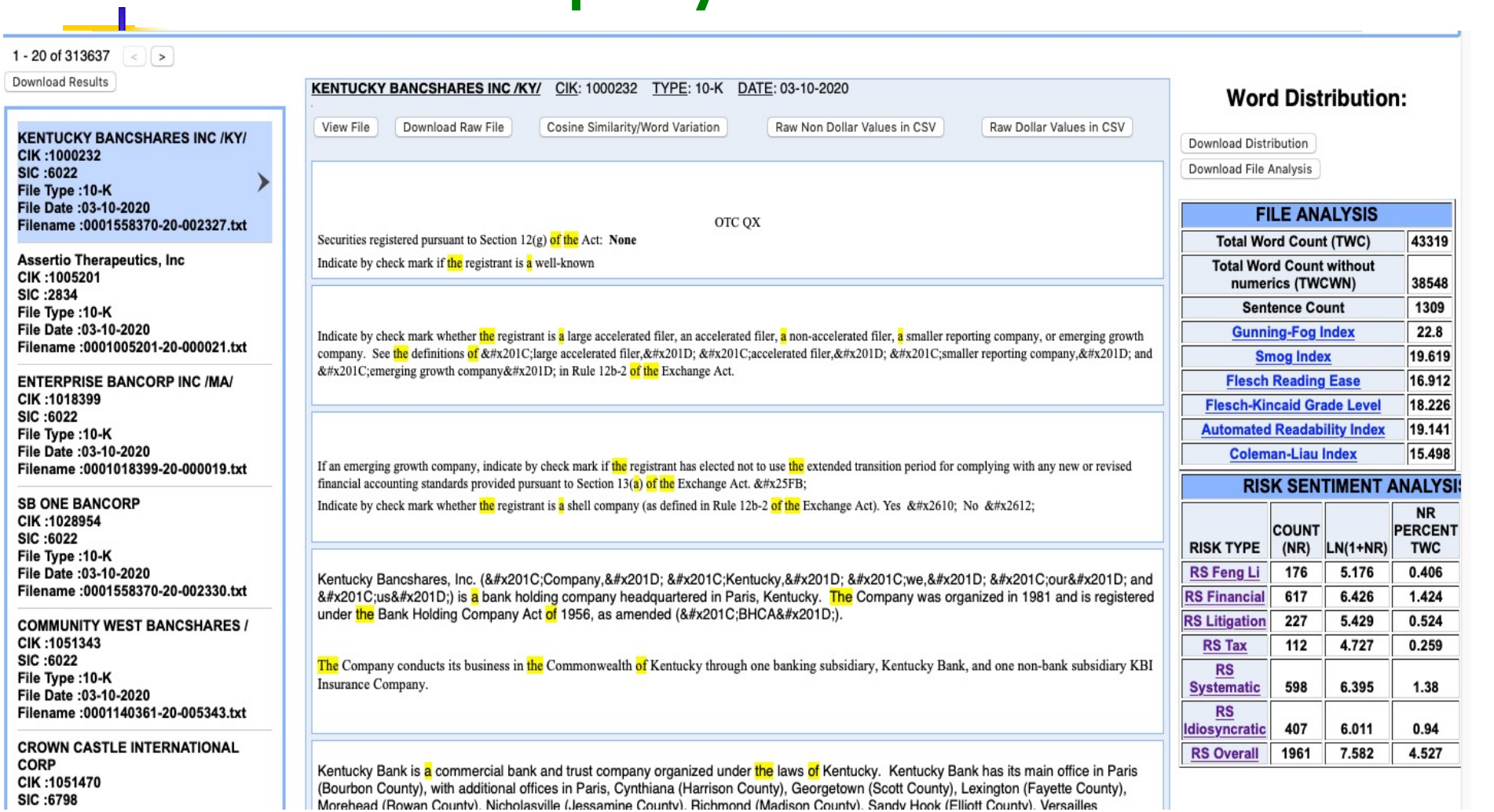
*** These extraction are from various filings, filed by companies to the SEC website and not from company websites.**

**** AAER & PCAOB do not have CIK's, they have File number and Firm number respectively**

STEP 5: SUBMIT

SeekEdgar

Display Window



SeekEdgar

Output in CSV File

CIK	COMPANY NAME	DATE	FILING TYPE	SECLINK	TOTAL WORD COUNT (TWC)	TOTAL WORD COUNT WITHOUT NUMERICS (TWCWN)	SENTENCE COUNT	GUNNING-FOG INDEX	SMOG INDEX	FLESCH READING EASE	FLESCH-KINCAID GRADE LEVEL	AUTOMATED READABILITY INDEX	COLEMAN-LIAU INDEX
1000232	KENTUCKY BANCSHARES INC /KY/	3/10/20	10-K	https://www.sec.gov/Archives/edgar/data/1000232/000155837 0-20-002327-index.html	43319	38548	1309	22.8	19.619	16.912	18.226	19.141	15.498

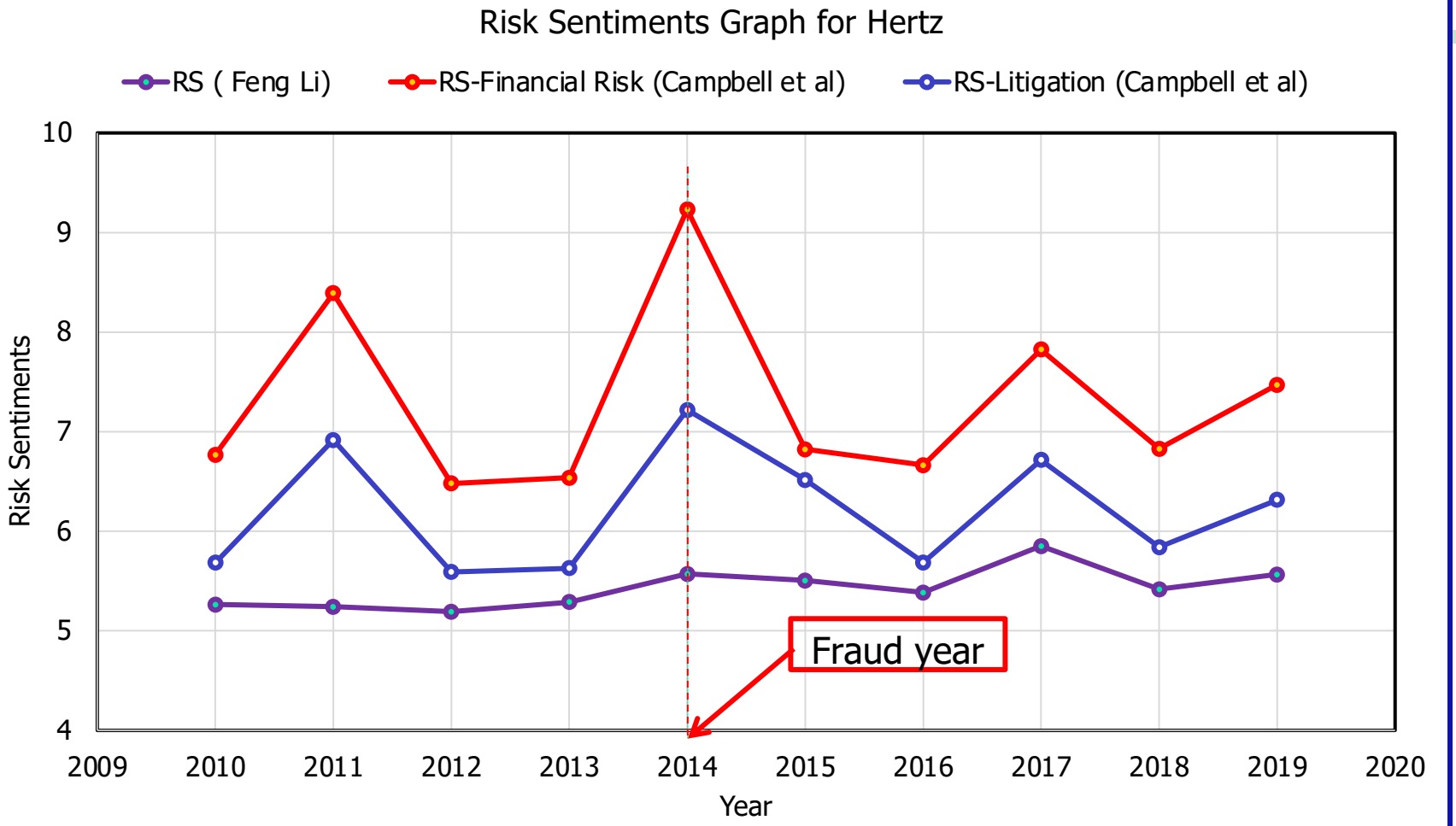
RS FENG LI - COUNT (NR)	RS FINANCIAL - COUNT (NR)	RS LITIGATION - COUNT (NR)	RS TAX - COUNT (NR)	RS SYSTEMATIC - COUNT (NR)	RS IDIOSYNCRATIC - COUNT (NR)	RS OVERALL - COUNT (NR)	RS FENG LI - LN(1+NR)	RS FINANCIAL - LN(1+NR)	RS LITIGATION - LN(1+NR)	RS TAX - LN(1+NR)	RS SYSTEMATIC - LN(1+NR)	RS IDIOSYNCRATIC - LN(1+NR)	RS OVERALL - LN(1+NR)
176	617	227	112	598	407	1961	5.176	6.426	5.429	4.727	6.395	6.011	7.582

RS FENG LI - NR PERCENT TWC	RS FINANCIAL - NR PERCENT TWC	RS LITIGATION - NR PERCENT TWC	RS TAX - NR PERCENT TWC	RS SYSTEMATIC - NR PERCENT TWC	RS IDIOSYNCRATIC - NR PERCENT TWC	RS OVERALL - NR PERCENT TWC	RS FENG LI - NR PERCENT TWCWN	RS FINANCIAL - NR PERCENT TWCWN	RS LITIGATION - NR PERCENT TWCWN	RS TAX - NR PERCENT TWCWN	RS SYSTEMATIC - NR PERCENT TWCWN	RS IDIOSYNCRATIC - NR PERCENT TWCWN	RS OVERALL - NR PERCENT TWCWN
0.406	1.424	0.524	0.259	1.38	0.94	4.527	0.457	1.601	0.589	0.291	1.551	1.056	5.087

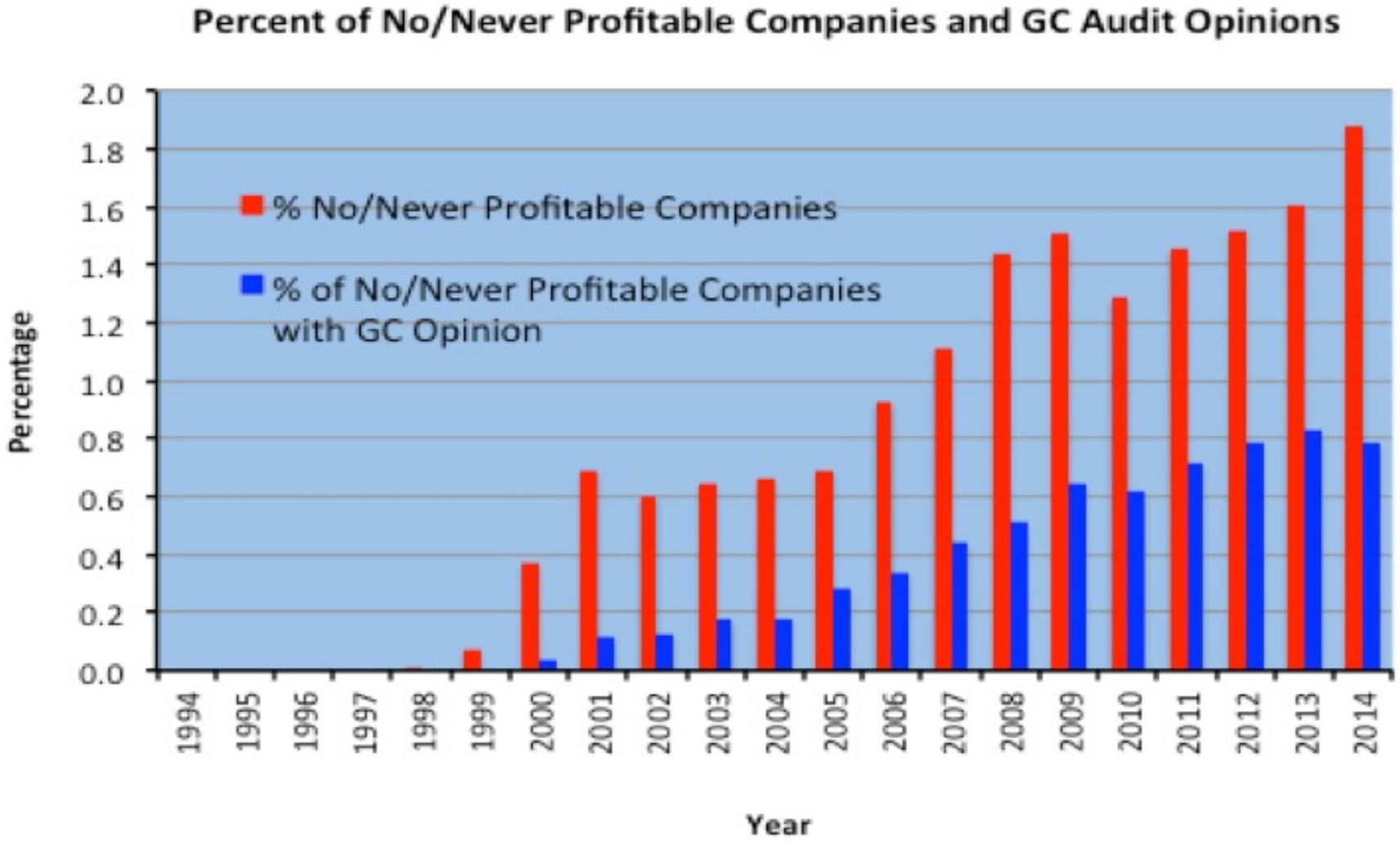
Seven Risk Sentiments in SeekiNF

- 10Ks, 10Qs, and their amendments (paragraph, all)
- 10Ks, 10Qs, and their amendments (Footnotes)
- 10Ks, 10Qs, and their amendments (MD&A)
- Item 1A of 10K (paragraph, all)
- 20Fs and 40Fs, and their amendments (paragraph)

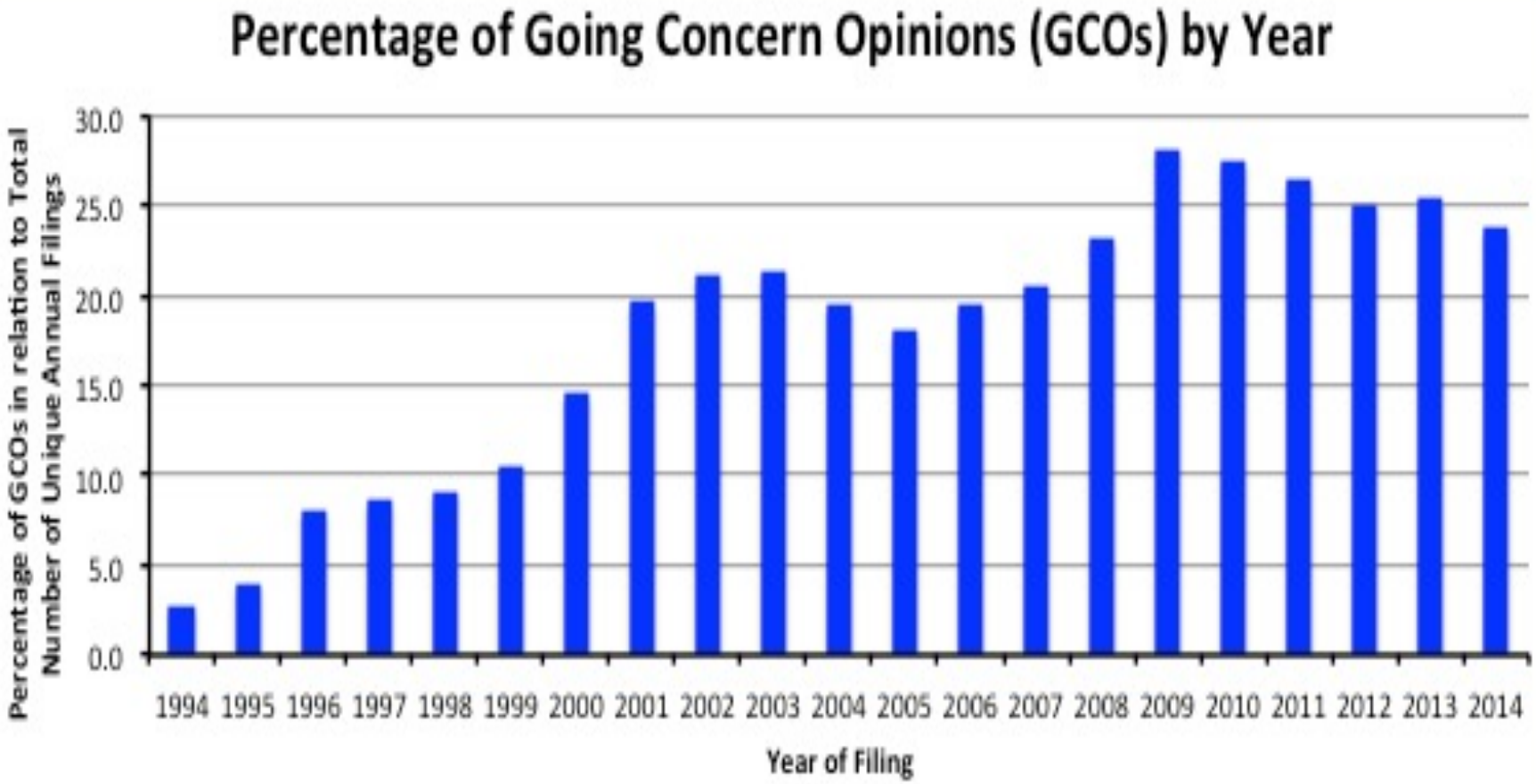
Risk Sentiments for Hertz Based on 10K

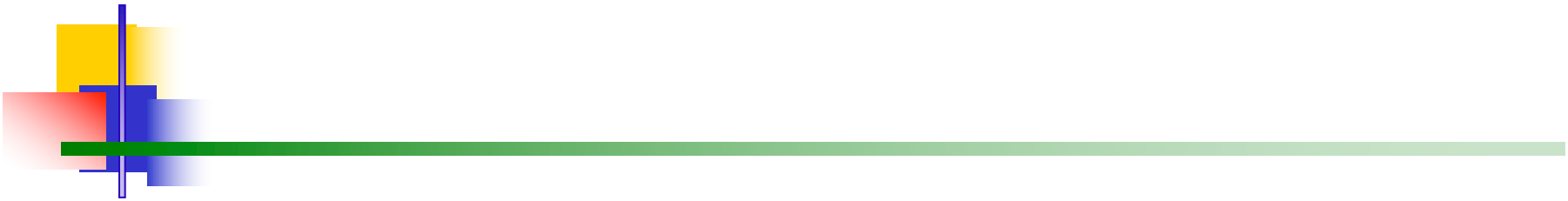


"We may never become profitable" in 10K with Going Concern Opinion by Year



Example: Graph of Going Concern Opinions by Year





Measure of Competition

Measure of Competition

Li, Lundholm, and Minnis *JAR*, 2013, p. 399

Li, Lundholm, and Minnis (2013) develop a model to compute management's perception of the intensity of competition using textual analysis of firms' 10-K filings.

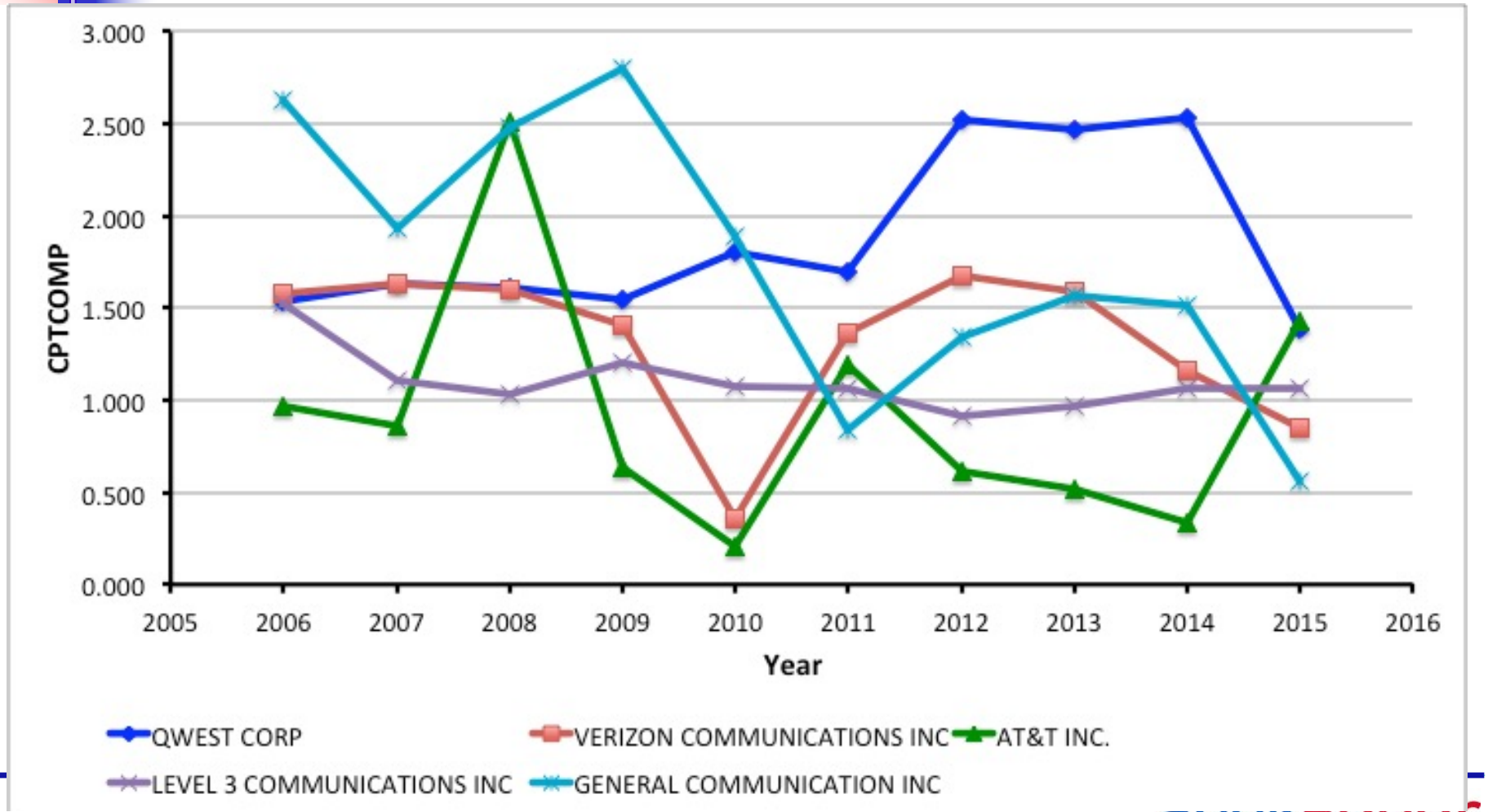
- Measure of competition varies across-industry and within-industry
- It is related to the firm's future rates of diminishing marginal returns.
- This measure is based on the count of the number of words like "competition, competitor, competitive, compete, competing," including those words with an "s" appended, less any case where "not," "less," "few," or "limited" precedes the word by three or fewer words.

$$PCTCOMP = 1000*NCOMP/NWORDS$$

where NCOMP = number of words in 10-K as described above and NWORDS = Total number of words without numbers.

SeekiNF SM

Competition Metric for Five companies for 10 years



Text Mining: Fraud Risk Assessment Model using Nonfinancial Measures

Brazel, Jones, and Zimbelman (*JAR,* December 2009)

Del Global Technologies (1997, Fraud)

Income: Overstated	$3.7 million.
Revenue:	25% from PY.
Employees:	6% (440 to 412)
Distribution Dealers:	38% (400 to 250)

Fischer Imaging Corp (1997, No Fraud):

Revenue:	27%
Employees:	20%
Distribution Dealers:	7%

Financial and Non-Financial Information from a Line

E32 | 11/5/1999

	A	B	C	D	E	F	G	H	I	J
1	COMPANY N.	CIK	SIC	FILE TYPE	FILING DATE	VALUES				
2	DEL GLOBAL	27748	3679	10-K405	11/13/96					
3										
4										
5	DEL GLOBAL	27748	3679	10-K405	11/13/96	The Company is a key supplier of noise suppression products for use i	400			
6	DEL GLOBAL	27748	3679	10-K405	11/13/96	Medical imaging systems dealers are supported by the Company's regional managers, product line managers and techni				
7	DEL GLOBAL	27748	3679	10-K405	11/13/96	Technical support in the selection, use and maintenance of the Company's products is provided to dealers and professio				
8	DEL GLOBAL	27748	3679	10-K405	11/13/96	The Company also maintains telephone hotlines to provide technical assistance to dealers and professionals				
9										
10	COMPANY N.	CIK	SIC	FILE TYPE	FILING DATE	VALUES				
11	DEL GLOBAL	27748	3679	10-K	11/14/97					
12										
13										
14	DEL GLOBAL	27748	3679	10-K	11/14/97	Marketing, Sales and Distribution The Company's medical imaging sys	250			
15	DEL GLOBAL	27748	3679	10-K	11/14/97	Medical imaging systems dealers are supported by the Company's regional managers, product line managers and techni				
16	DEL GLOBAL	27748	3679	10-K	11/14/97	Technical support in the selection, use and maintenance of the Company's products is provided to dealers and professio				
17	DEL GLOBAL	27748	3679	10-K	11/14/97	The Company also maintains telephone hotlines to provide technical assistance to dealers and professionals				
18										
19	COMPANY N.	CIK	SIC	FILE TYPE	FILING DATE	VALUES				
20	DEL GLOBAL	27748	3679	10-K	11/12/98					
21										
22										
23	DEL GLOBAL	27748	3679	10-K	11/12/98	Marketing, Sales and Distribution The Company's medical imaging sys	250			
24	DEL GLOBAL	27748	3679	10-K	11/12/98	Medical imaging systems dealers are supported by the Company's regional managers, product line managers and techni				

1996-DelG +

Corporate Governance
Effectiveness of Board of Directors
Interesting Issues

- Independence
  - CEO and Chairman of the BOD
- Competence
  - No Financial Expertise in the Board or Audit Committee
- Activity
  - Number of Meeting of BOD, Audit Committee, and Risk Committee

Unique Databases Available

Corporate Governance Related databases:

1. **CEO is Chairman of the BOD** for 1994-2019.
2. **No Financial Expertise in BOD** for 1994-2019.
3. **Number of Audit Committee Meetings** for 1994-2018.
4. **Number of Board of Directors Meetings** for 1994-2018.
5. **Number of Risk Committee Meetings** for 1994-2018.

Other Databases:

6. **Compensation and Executive Bio** from DEF 14A for 2003-2020.
7. **Compensation and Executive Bios** from Form 424Bs, 2003-2020.
8. **Public Float Value** from Annual Reports, 1994-2020.
9. **Accounting and Finance Executives who Signed 10Ks**
10. **Subsidiaries database** by year and country (1994-2020).

Unique Databases Available

11. **Conference calls** (who said what?), 1994-2020, from SEC Filings.
12. **Company specific information** for the years 1994-2018 from 10Ks such as: CIK, Company Name, Business Address, Mailing Address, Filing Date, Date Accepted, Period of Report, IRS No., State of Incorporation, Fiscal Year End.
13. **File Analyses for all 10Ks and 10Qs for 25 years (1994-2018)**: Word Count, Word Count without Numeric, Total Number of Sentences, and six readability indices.
14. **Database based on Forms 3, and 4.**
15. **Competition Metrics** for all companies based on 10K Filings (1994-2020).
16. **Companies with Going Concern Opinion,** 1994-2020

Additional Databases That One Can Create

In addition to the previously listed databases, one can create, in a matter of seconds, the following databases:

1. **All Audit Opinion types** such as **clean opinion** and **going concern opinion**
2. **Audit Fees** from DEF 14A
3. **Which Institutional Investor invests where** through N-Q and N-CSR
4. many more such databases. Let me know if you or your colleagues would like to know how to create such databases using SeekiNF.

2019-2020 Subscribers

1. Arizona State University
2. Australian National University, Canberra, AU
3. Bentley University, Boston
4. Brock University, Canada
5. City University of Hong Kong
6. Fordham University,
7. IIM-Ahmadabad, India
8. Louisiana Tech University
9. Macquarie University, Sydney, AU
10. McMaster University, Canada
11. Nanyang Technological Univ, Singapore
12. National Central University, Taiwan
13. Rutgers University – Newark
14. Southern Illinois University
15. University of Alabama–Huntsville

- University of Arkansas
- **University of Chicago**
- University of Illinois-Chicago
- University of Louisville
- University of Kansas
- University of Nebraska - Lincoln
- University of Queensland, Australia
- University of Southern California
- University of Sydney, Sydney, AU
- University of Texas-San Antonio
- Univ. of Waterloo, Canada
- Villanova University
- **Washington Univ., St. Louis**
- Xavier University
- **Yale University**

SeekiNF SM

SeekEdgar

Anytime, Anywhere...
SEC filings at your fingertips in seconds with SeekiNF

Webinar on how to setup search criteria in SeekiNF, two slots, every Wednesday.
Register HERE for the webinar

SeekiNF

SeekiNF(Search Engine to Extract Knowledge from Industry Filings) is a cloud-based technology that provides searchable interface within different levels of resolutions in SEC filings. SeekiNF allows users to search for financial and non-financial information based on user specified search strings from the various SEC filings and PCAOB Inspection Reports from 1994-2019 with daily updates. It has around 17 million SEC filings and 33 million documents in its database spread over 24 years. SeekiNF provides tremendous opportunities for gathering information related to accounting, finance, marketing, tax, law, etc. to conduct research and develop various risk assessment models including financial risk and fraud risk using financial and non-financial information. In short, **SeekiNF is a Complete System, Second to None with its own Database and Incredible Search Engine**.

FRAANK

FRAANK allows users to get financial statements, especially Balance Sheet, Income Statements, and Cash Flow Statements, of public companies filed with the SEC. These statements are made available in Excel Spreadsheet format.

Also, FRAANK lets users download non-financial information from 10K and 10Q filings, such as: Audit Report, Summary of Audit Report - auditee, auditor, type of opinion, place and date of audit, and dumps into Excel Spreadsheet, SOX 404 Management and Audit Reports, Footnotes, Items, and MD&A, Signature - obtains names of people who signed 10K with their titles.

SeekINF SM

Anytime, Anywhere!

SEC Filings & PCAOB Reports

at

Your Fingertips

in Seconds

SeekEdgar

Anytime, Anywhere, at Your Fingertips!

SeekiNF℠

Access SeekiNF using

- iPad
- Smart Phone
- Laptop
- Desktop

Anytime, Anywhere, at Your Fingertips!

SeekINF SM

19 million Filings

and

35 million Documents

(1994-2021)

Daily Updated

Anytime, Anywhere, at Your Fingertips!

SeekiNF SM

A Complete System! Second to None!!

With its own Database

&

Incredible Search Engine.

You have to try it to believe it!

SeekEdgar

Anytime, Anywhere, at Your Fingertips!

SeekINF SM

Search All Filings by

- Paragraphs (All)
- Tables & Footnotes
- Audit Reports
- SOX 404 Reports
- MD&A

Anytime, Anywhere, at Your Fingertips!

SeekINF SM

Search All Filings by

- CIK Codes
- SIC Codes
- Ticker Symbols

Anytime, Anywhere, at Your Fingertips!

SeekINF℠

Search by

- One Company
- Multiple Companies
- All Companies
- One Filing
- All Filings

SeekINF SM

Unique Features for Textual Analysis

- **Word count**, **Word count without numbers**, **Sentence count**, available for all documents
- **Proximity counter** with Order and without Order, available for all documents
- **Word Distribution**, available for all documents
- **Six Readability Indices**, available for all documents
- **Risk Sentiments metrics** (Financial Risk, Litigation Risk, Tax Risk, Idiosyncratic Risk, System Risk, and Overall Risk) for 10Ks, and 10Qs, Item 1A, MD&A and Footnotes in 10Ks and 10Qs.
- **Cosine Measure of Similarity**, within company and for one company versus another, available for all the SEC filings and PCAOB Reports. See the user's guide at https://www.seekedgar.com/UserGuide.pdf for interesting fraud related examples.
- **Word Variation,** year-to-year, all documents

SeekINF SM

Additional Unique Features

- **Use Boolean Logic** to search for information with multiple words/phrases with no Stop Words.
- **Obtain Unique financial information** such as Air Traffic Liability, Fuel Hedge Contracts, etc. in a matter of seconds.
- **Display few words before & few words after** a phrase.
- **Extract any number**, financial or non-financial, mentioned in a line in a document in Excel such as number of active patents.
- **Download Results** of all your searches including tables, and snippets in Excel and HTML documents.
- **Get the entire document** of a specific type by typing "a | the | of" without the quotes in Step 1.

SeekiNF SM

Unique Search Features

(Not Available from any Data Provider)

1. Boolean Logic (use + for AND, | for OR, - for Negation)

- Multiple Exact phrases **(with no stop words)**
- With all of the phrases/words **(use A + B + C)**
- With at least one of the phrases/words **(use A | B | C)**
- Without certain phrases/words **(A - B, with A but not B)**

- Examples (No hyphen, instead use space, e.g., risk-free as risk free)
- **Chairman of the Board of Directors, Chief Executive Officer and President** | Chairman of the Board of Directors and Chief Executive Officer | Chief Executive Officer and Chairman of the Board of Directors | President, CEO and Chairman of the Board of Directors | CEO and Chairman of the Board of Directors| Chairman of the Board and Chief Executive Officer | Chairman of the Board and Chief Executive Officer in **Table** in all 10K types, got 21,940 for 1994-2017, in Table in 10 Ks in 28 seconds.
- **We do not have a qualified financial expert** serving on our board of directors | We do not have an audit committee financial expert serving on our Board of Directors | No member of our Board of Directors qualifies as an audit committee financial expert in Exact Phrase, Paragraph, 2 minutes got 309 hits.

More Examples of Phrase Searches Using Boolean Logic

- **For Going Concern Opinions:** (We have audited | I have audited) + (substantial doubt about| substantial doubts about | substantial doubts regarding | substantial doubt regarding) + going concern (Try 1994 and 2018)
- **For Clean Opinion:** (We have audited | I have audited) + present fairly + in all material respects - (substantial doubt about| substantial doubts about | substantial doubts regarding | substantial doubt regarding) - going concern
- **Audit Fees:** audit fee | audit fees | audit related fees | audit related fee in Exact Phrase, Table, in DEF 14A
- **Phrase Counter:** Litigation and Litigation Counts, Try for 2019
- futures + options + swaps + hedging, Paragraph, in10K
- (block chain | blockchain | distributed ledger) - chain link

Unique Search Features
(Not Available from any Data Providers)

2. **Wild Card Search** (works only with all of the words feature, **Example:** Risk*)

3. **Proximity Search** (two or more words within few words)
   - **Compensation data** (Bonus Salary, within 2 words, table, DEF 14A)
   - **Executive bios** (Name age within 2 words, table in DEF 14A)
   - **Who signed 10Ks**, etc. (title signature within 4 words, table, 10Ks)
   - **Who invests where** (Uber Preferred within 10 words in Table, All)
   - Infosys shares value within 300 words in Table in N-CSR and NQ (Institutional investors)
   - Who all institutional investors have invested in General Motors, Exact Phrase: General Motors, Proximity: shares value within 2 words, Table, N-CSR and N-Q.

Unique Search Features
(Not Available from any Data Provider)

4. Display (few words before & after a phrase)
   - Coronavirus
   - Fiscal Year End from 10K (0, 1)
   - STATE OF INCORPORATION, 10K (0, 1)
   - Accession Number, 10K (1, 0)
   - Business addresses, 10K (0, 22)
   - audit committee met in DEF 14A (0, 1, try 0, 5)

Obtain Any Financial and Non-Financial Data

- Off Balance Sheet Items
- Air Traffic Liability
- Fuel Hedge Contracts
- Advertising expenses
- Marketing expense
- Off Balance Sheet items
- R&D Tax Credit (use R D Tax Credit)
- Executive Compensation Tables
- Executive Bios

Liu and Moffitt

(*Journal of Emerging Technology in Accounting, 2016*)

- Textual analysis of SEC Comments Letters and developed a measure of intensity based on the modality of comment letters.
- Observed that the intensity of comment letters is positively associated with the probability of a restatement of the reviewed 10-K filings.
- Moreover, textual analysis and text mining techniques provide information about companies' performance that is not available otherwise.

Assessment of Financial Risk and Fraud Risk using Textual Analysis

- "Detect Fraud Before Catastrophe" by Lee, Churyk, and Clinton, Strategic Finance, March 2013, p. 33.
  - Proactive content analysis techniques can help management accountants prevent catastrophic financial fallout.
- "Using Nonfinancial Measures to Assess Fraud Risk" by Brazel, Jones, and Zimbelman, JAR 2009, p. 1135.
- SEC: Corporate Filers Beware: New "RoboCop" is On Patrol
  - Based on AQM and Text Analytics (not used yet, some companies are working on it)

Fraud Risk Assessment Model using Textual Analysis

Lee, Churyk and Clinton (*Strategic Finance* , 2013, p. 33)

Fraud detection model based on the textual, i.e., content, analysis of MD&A in 10-K:

$$Fraud_i = 2.89757 - 0.83408\ (Positive\ Emmotion_i)$$
$$- 0.48315\ (Present\ Tense_i)$$
$$+ .0001\ (Total\ Words_i)$$
$$- 2.80753(Colons_i)$$

"Conventional fraud detection measures using ratio analysis and other financial data were either unable to detect the fraud or unable to detect it soon enough to avoid catastrophic outcomes".

Professor Michael Minnis
University of Chicago

"I've really appreciated working with SeekEdgar.

The flexibility and speed of both the search engine and customer service has **made data collection easier and more thorough**.

Finding examples for my financial statement analysis class is also a pleasure.

I recommend trying SeekEdgar for anybody looking for data from SEC filings

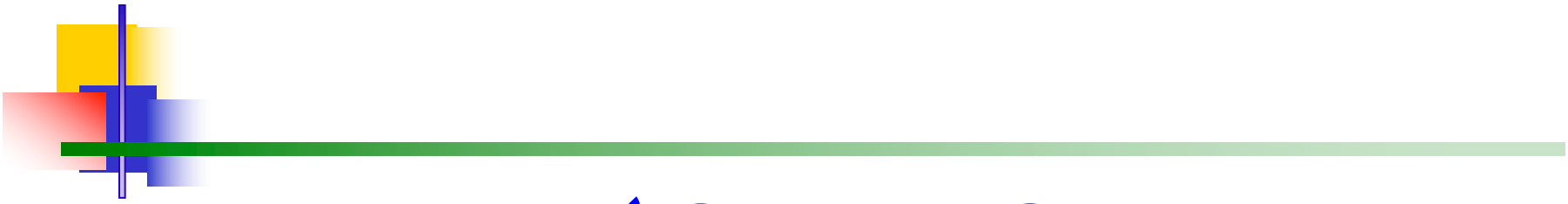
Conclusion

"Knowledge, Imagination, and Creativity

will drive the future research,

not the canned data"

Dr. Rajendra P. Srivastava
rsrivastava@seekedgar.com
rsrivastava@ku.edu



Thanks!!

Questions?

SeekEdgar