

If you had everything computationally...
...where would you put it, financially?

David Leinweber
Haas Fellow in Finance,
Haas School of Business
U.C. Berkeley
djl@haas.berkeley.edu
dleinweber@post.harvard.edu



c. 2002-2008, D. Leinweber

Talk at O'Reilly Money:Tech, Feb. 6, 2008

Why am I here?

IAC / ITG QuantEx

- Quantex Algos
- Jefferies Acquisition, ITG Spinoff



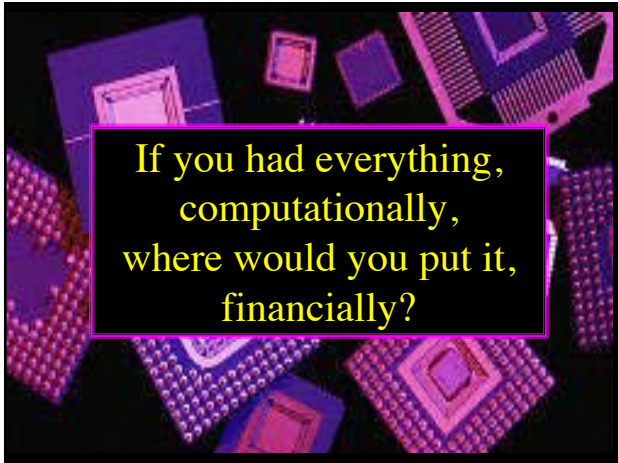
- MD for Equities
- Institutional Buy side
- \$6 Billion
- 6 countries, 27 quant strategies, Long & MN



- Internet Information Service, Founder
- Integrating textual information in trading strategies
- Caltech, post bubble



- Center for Innovative Financial Technology



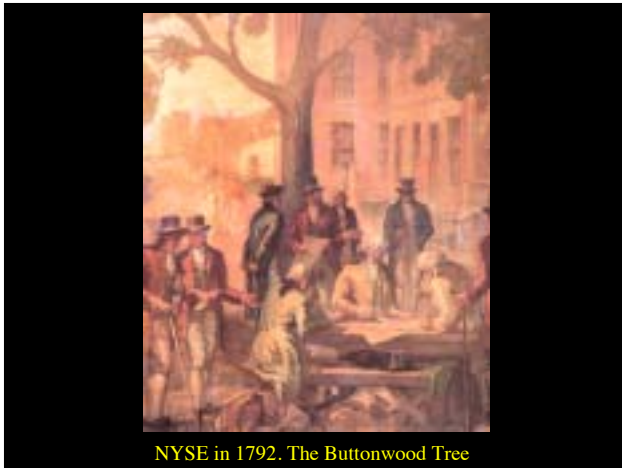
Summary of JPM Article

- **Looking Back: Greatest Financial Technology Hits**
 - Electronic Market Data
 - Computerized Data Storage and Analysis
 - Electronic Execution
- **Looking Forward: Past as Prelude**
 - Algos to the Nth Power
 - Intelligence Amplification and Visualization
 - **Finding Alpha in Textual and Internet Information**



Journal of Portfolio Management Fall 2005, pp 61-75

A Short History of
Market Information
Technology



NYSE in 1792. The Buttonwood Tree



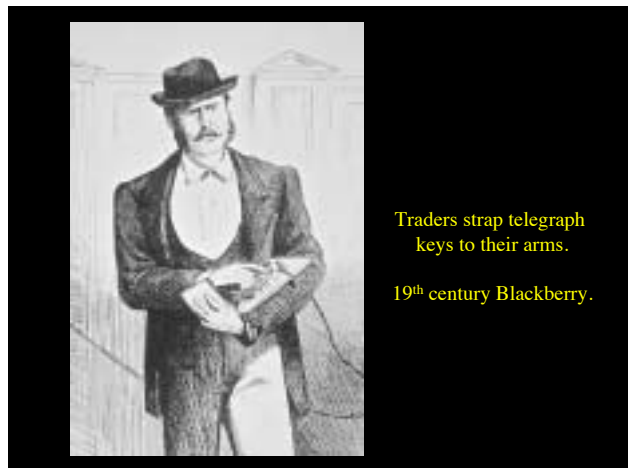
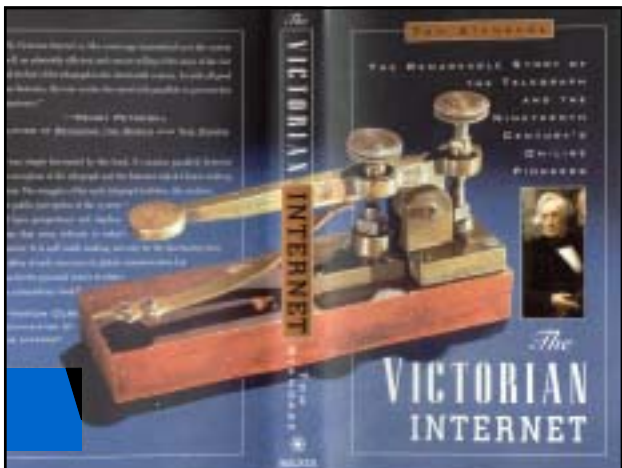
NYSE moves indoors. Tontine Coffee House. 1794



THE WONDER of the AGE !!
 THE SALVIFIC AND ELECTRO-MAGNETIC
TELEGRAPHS,
 OF THE WESTERN RAILWAY.

The Electric Fluid travels at the rate of 200,000 Miles per Second.

ADMISSION ONE SHILLING.



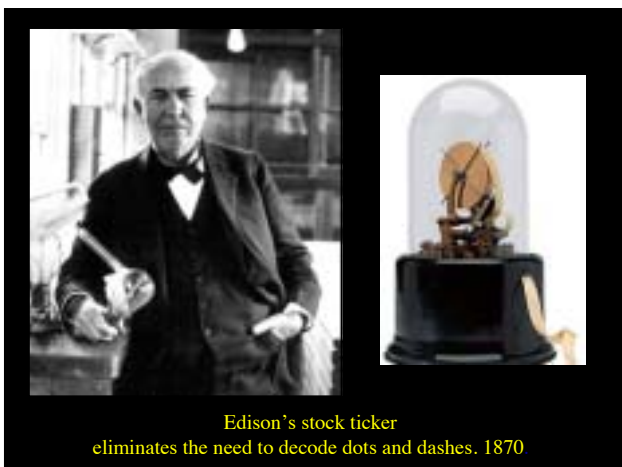
Traders strap telegraph keys to their arms.
19th century Blackberry.



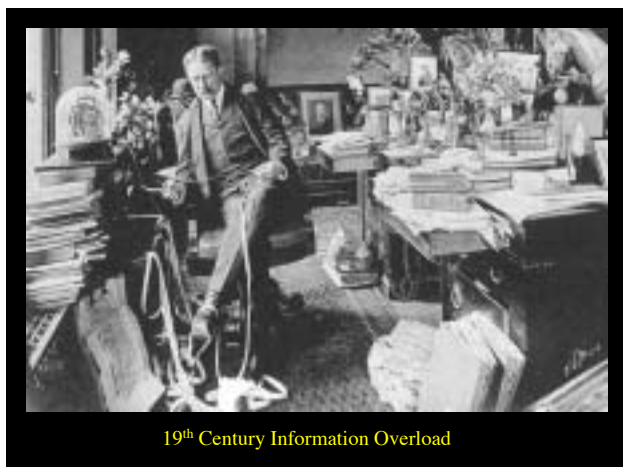
NYSE Before Telegraphy, c. 1865



Telegraph wires at the NYSE, 1888.



Edison's stock ticker eliminates the need to decode dots and dashes. 1870



19th Century Information Overload



NYSE Quote Board. 1930-40s



Market Data Archive c. 1950s



Telephone Order Overload, Dealing Room, c.1950, Reuters



NYSE Floor - 1963

NYSE Photo



Beyond Ticker Tape

Exchange President
Keith Funston with
first NYSE computer.

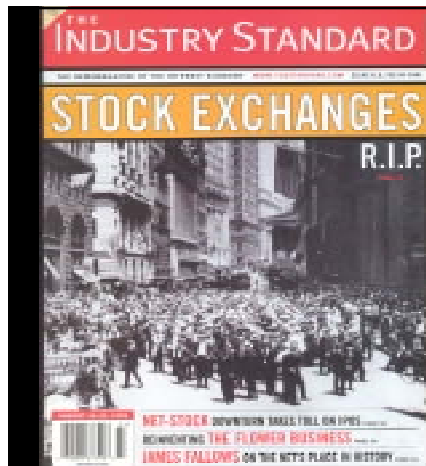
1966

Disintermediation of Execution



Electronic
Disintermediation
of Execution

E*trade, 1999



An Internet-centric
trade journal buried
the traditional exchange
in a 1999 issue.

The exchanges are
still here. Changing,
but here.

The Industry Standard
isn't.

(Aug. 16, 1999)



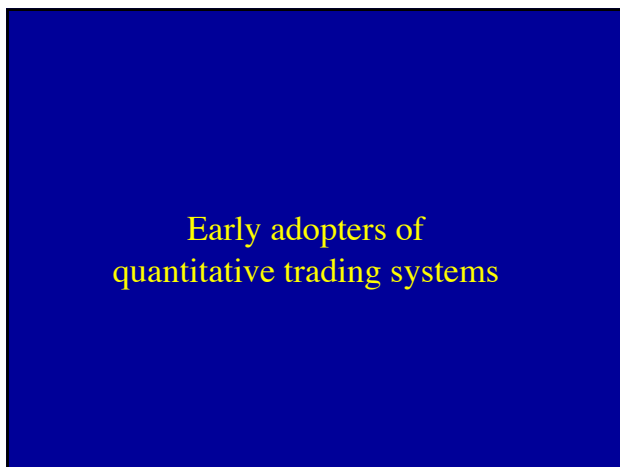
London Stock Exchange. Big Bang minus 1. Oct 26, 1986



London Stock Exchange. Big Bang. Oct 27, 1986



Early adopters of quantitative trading systems



Fischer Black: Options Maven and Pioneer Algo Trader

"Markets look much more efficient from the banks of the Charles than from the banks of the Hudson."



**Renaissance Technologies
Founded 1982 and still not talking**



Information Advantage: 1980s-90s *Algorithmic Market Making & Trading*



"UNIX and Market Data Feeds" ~ 1986

Fortune, Feb. 5, 1996

Information Advantage: 1990s + *Algorithmic Market Making*



"Playing NASDAQ like a piano" ~ Dave Whitcomb

What's Next?

- Exploitation of electronically delivered quantitative information was, and remains, a great success.
- Prices and market data convey a great deal of information. But they do not convey *all information*.

Where to look for new information advantages?



Where does alpha come from?

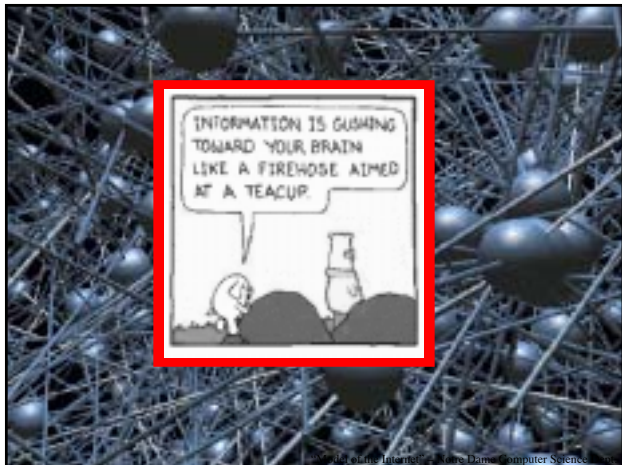
Information & Innovation

"...profits may be viewed as the economic rents which accrue to [the] competitive advantage of... superior information, superior technology, financial innovation..."

Andrew W. Lo, MIT Sloan School

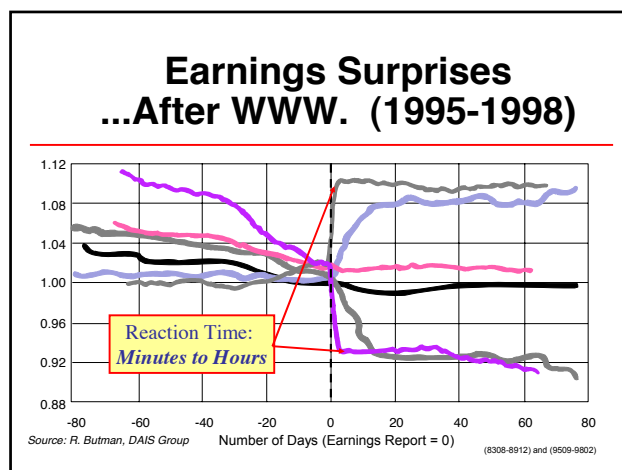
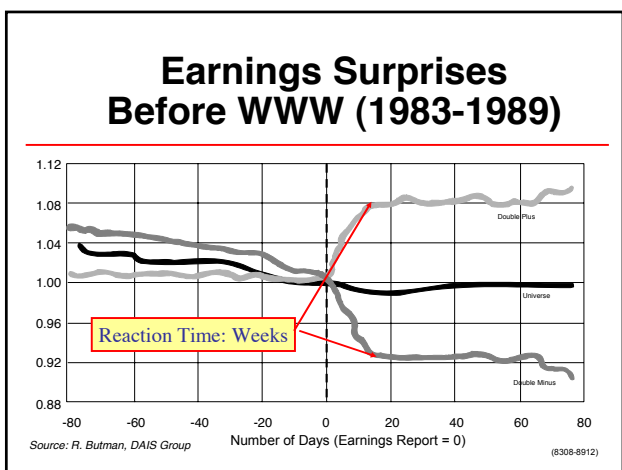
Editor, "Market Efficiency: Stock Market Behavior in Theory and Practice", Elgar, 1997.



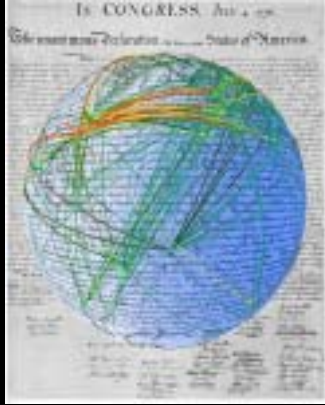


Firehose of Information With Potential Alpha


- Electronic versions of mainstream news
 - Minutely instead of Monthly
 - "Print", Web, Broadcast captions, feeds.
 - Global news on global companies. Time zone advantage.
- Press releases, disintermediated access
 - PRNewswire, BusinessWire
 - Specialized sector news
- Electronic access to "official sources"
 - SEC, Courts, NIH, federal & state agencies
 - Management conference calls
- New Media – Creators of the net
 - Websites, mail, messages, chat, blogs, RSS...
 - "Social Media"



The Text Frontier: *Finding Alpha on the Web*



**Democratization
and
Disintermediation
of
Information**



Ace
Reporters?

Who needs 'em!

Do it Yourself

Disintermediated
News



**Does news
move
markets?**

Off the shelf algo news products

- 2007 Product Announcements
 - Dow Jones
 - Reuters
- Acquisitions & Strategic Partnerships
 - Clearforest/Reuters
 - Corpora
 - Ravenpack
- Not cheap at all
 - ~ \$100K per month
- Can one size fit all
 - Does it get arbed away rapidly?



REUTERS
Professional products

Readers for Algorithmic Trading

Programmatic, "start up", "model-driven", "black box", whatever you call it, algorithmic trading is reshaping the securities industry.

Readers for Algorithmic Trading

Readers for Algorithmic Trading

eAnalyst Case Study

News, Language Models,
One day horizon

November 13, 2001

Plugged In

Can Computerized Language Analysis Predict the Market?

Mining of Concurrent Text and Time Series

Victor Lavrenko, Matt Schaeff, Dawn Lawrie, Paul Ogilvie,
David Jensen, and James Allan
Department of Computer Science
University of Massachusetts
Amherst, MA 01003

UMass Working Paper, 2001

Source: Victor Lavrenko

Analyst: Overview

Language Models on Stock News

- Estimate a language model for each trend
 - use trends instead of raw data
 - align trends with concurrent news stories
- Given a new document, predict which trend will most likely occur next

Source: Victor Lavrenko

Extracting Trends

- Replace series by a sequence of regression lines
 - top-down procedure (slide a window, split, recurse)
 - automatic stopping criterion based on T-test

Source: Victor Lavrenko

Language Modeling

Microsoft (MSFT) stock

News: [Icons of news stories]

Words like **Jackson** and **antitrust** are more likely in the stories preceding the plunge.

Software giant Microsoft saw its shares dip a few percentage points this morning after U.S. District Judge Thomas Penfield Jackson issued his "findings of fact" in the government's ongoing antitrust case against the Seattle wealth-creation machine...

$P(\text{shares}) = 0.074$ $P(\text{antitrust}) = 0.009$ $P(\text{judge}) = 0.006$ $P(\text{trading}) = 0.032$ $P(\text{against}) = 0.025$ $P(\text{Jackson}) = 0.001$	$P(\text{shares} \text{MSFT}\downarrow) = 0.071$ $P(\text{antitrust} \text{MSFT}\downarrow) = 0.044$ $P(\text{judge} \text{MSFT}\downarrow) = 0.039$ $P(\text{trading} \text{MSFT}\downarrow) = 0.029$ $P(\text{against} \text{MSFT}\downarrow) = 0.027$ $P(\text{Jackson} \text{MSFT}\downarrow) = 0.025$
---	---

$P(\text{MSFT}\downarrow | \text{Jackson}) = P(\text{Jackson} | \text{MSFT}\downarrow) P(\text{MSFT}\downarrow) / P(\text{Jackson})$

eAnalyst Evaluation: Trading Simulation

- Cumulative earnings: \$21,000
 - 40 day simulation, out of sample
 - 100 stocks, Jan 1, 2000
 - purchased or shorted \$10,000 with each trade
- Result significant at 1% level
 - determined through a randomization test
- Biggest Gainers
 - IBM: \$47,000
 - Lucent: \$20,000
- Biggest Losers
 - Disney: -\$53,000
 - AOL: -\$18,000

Source: Victor Lavrenko

Tetlock's Results

Predictability & Consistency

The main result in Table III is that negative words in firm-specific news stories strongly predict lower returns on the following trading day. The coefficients on negative words (neg) are strongly negatively significant in all 12 of the regressions when news stories from DUVS are included.

Alpha: 21.1% cumulative annual return (1984-2004)

We hold both the long and short portfolios for one full trading day and rebalance at the end of the next trading day. To keep the strategy simple, we exclude the two days in which either the long or the short portfolio contains no qualifying firms. Ignoring trading costs, the annualized cumulative returns of the long-short strategy would be 21.1%.

Disintermediated News Primary Sources Available on the Net

- Specialized Industry Media
- Local and International Media
- Direct Corporate Communications
- RIXML and other research
- Government Agencies
 - Courts
 - Regulators
 - SEC

The SEC Mother Lode



- Reputable
- Large, Broad
- Deep Web
 - In databases
- Tractable
 - Real use of XML

Speech by SEC Chairman
The Honorable Christopher Cox

Joint International XML Conference
Philadelphia, Pennsylvania
November 5, 2006



Deep web – Where's the SEC?



Deep, and Semantic Too



SEC Chairman Christopher Cox outlines his plan to increase regulatory transparency by making financial reports available in XML.

SEC gets the XML Religion

Cox will request the plug for XBRL, the Extensible Business Reporting Language, a means of tagging financial data that, he said, is a Spanish for Denmark, "tell us for business reporting what bar coding did for product distribution." On Tuesday, Cox reaffirmed his belief that reports written in plain English, and incorporating XBRL, would be "more meaningful and intelligible to the average investor."



Copyright ©
Commission Data and XML Initiative

Plain Old Ugly Edgar

```

<SEC-DOCUMENT>
<SEC-HEADER>
<SECURITY-CERTIFICATE>
<SECURITY-CLASSIFICATION>
<STATUS OF REPORT>
<DATE OF CHANGE>
<FILE NUMBER>
<PUBLIC DOCUMENT NUMBER>
<COMMISSION FILING OFFICE>
<FILING DATE>
<FILING TIME>
<SEC ACT>
<SEC FILE NUMBER>
<FILE NUMBER>
<INDEXED ADDRESS>
<OFFICE>
<CITY>
<STATE>
<ZIP>
    
```

New XML Hip Edgar

Multilevel Traffic Analysis of SEC Filings

Notorious Example #1

“What’s news?” - Aggregation & Cross-referencing of SEC filings

- Across multiple filing types
- Across multiple filing entities
- A multi step process
 - Find 10K section 21
 - Extract potential filing entities
 - Get all filings for all entities



Filings & Filers Analysis

Enron example

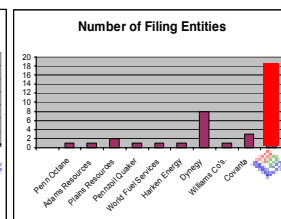
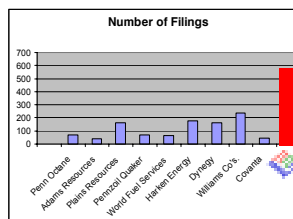
- How do the number and type of filings made by Enron and associated entities compare with other comparable firms?
- Historical footnote:
 - Traffic Analysis is the oldest and simplest form of electronic surveillance, dating back to WWII.

Methodology

- Find companies comparable to Enron
 - Same detailed SIC code: 5172
 - Wholesale-Petroleum and Petroleum Products
 - Or, Same detailed S&P Industry Group: 55103010
 - Multi-Utilities
- Look up all electronic Edgar filings associated with that company name on www.sec.gov
- Collect **number of filings** & **number of filing entities**
 - Measures of complexity and potential obfuscation

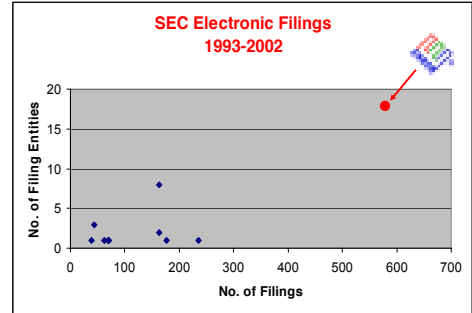
Striking Observation

- **Enron is a massive outlier on both scales**
 - Triple the average number of filings: **576 vs. 160**
 - Five times the average number of filing entities: **18 vs. 3.7**



A Barrage of...

Company	Number of Filings	Number of Filing Entities
Penn Octane	69	1
Adams Resources	39	1
Plains Resources	163	2
Pennzoil Quaker	70	1
World Fuel Services	63	1
Harken Energy	177	1
Dynegy	164	8
Williams Co's.	235	1
Covanta	44	3
Enron	576	18
Average	160.0	3.7
Std Dev.	160.9	5.5



Diffing Financial Footnotes In Successive SEC 10Q Filings

Leinweber & Sisk, 2003

Diff'ing Financial Footnotes

- Footnotes to financial statements in 10Q and 10K filings tend to remain the same from filing to filing
- When there are substantial changes, additions, or deletions, they are often interesting
- Diff'ing is a common programmer's tool for tracking changes in code
- If we extract the footnotes (hard, but getting easier), and diff, maybe we find something interesting

Symbol: FMC Form: 10-Q

From 11/14/2000 To 05/15/2001 (Page 20 out of 22 Representing 08/12/1994 to 08/14/2001)

SEC Form 10-Q for FMC (FMC CORP) for the period ending 08/14/2001. The document contains financial statements and footnotes. A red arrow points to a section of text in the footnotes.

Lot's of new discussion about derivatives



Symbol: HNWK Form: 10-Q
 From 05/15/2001 To 08/11/2001 (Pair 9 out of 18
 Representing 05/14/1998 to 08/11/2001)

An accounting tale of great complexity.



Symbol: PCG Form: 10-Q
 From 05/15/1997 To 08/14/1997 (Pair 0 out of 15
 Representing 08/14/1997 to 08/02/2000)

Energy trading.
 Additions are interesting.



Does better governance lead to better returns, as CALPERS and other large investors believe?
 Inferring Quality of Governance from SEC Filings

Director Board Membership Concentration and Returns

David Leinweber & Jacob Sisk
 Caltech / UCLA / Leinweber & Co.
 May 2006

Is better governance associated with better stock performance?

- Indicators of quality governance
 - Independence
 - Insider/Outsider ratio
 - Board concentration
 - How many other boards do directors sit on?
 - Turnover & Tenure
 - Stock ownership
 - No insiders on audit & compensation committees
- Much of this can be extracted from the sequence of 10K and 8K filings

Director Board Membership Concentration and Returns

- One component of quality measurement
- Based on conversations on potentially useful information in SEC materials, we investigated the relationship of board membership concentration and returns
- Board membership concentration =

$$\frac{\text{Total number of boards directors are on}}{\text{Number of directors}}$$

Hypothesis & Data

- Hypothesis
 - Over-extended boards, with directors who sit on many boards, will do a worse job for the company than boards that are more focused.
 - Lower concentrations should be associated with superior returns
- Data
 - Extractable from 10K, clean current data also in specialized web databases
 - We built a specialized spider to collect this data.

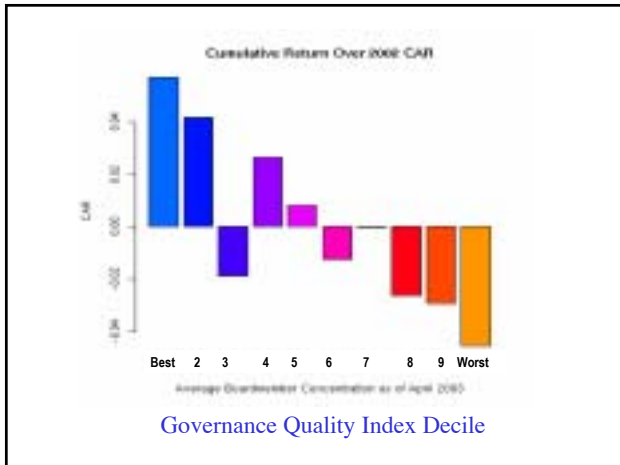
Methodological Notes

- Board concentrations are as of April 2003
- Group firms into quintiles & deciles by concentration
- Look at returns to each quintile/decile
 - Period is Jan. 1, 2002 – Oct. 31, 2002
- This is retrospective, along the lines of the Merrill Lynch size studies, but board concentrations tend to be stable.

Results

- Hypothesis is strongly indicated to be true.
 - “Textbook” results in both quintiles & deciles
- Excess return spreads (CY 2002)
 - 7% Quintile 1 – Quintile 5
- Large breadth
 - All companies have a rank
- Details follow





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

Feng Li, U. Michigan, 2006

Source: Feng Li 2006

Ranking MD&As

Li 2006

- How badly does the CEO want to not sound scared, unsure, worried?
- All firms filing with SEC, 1994-2005
 - Financials excluded, due to language
 - 34180 firm-years
- Extract Management Discussion & Analysis from 10K (Annual Report)
- Rate MD&A for **Risk Sentiment**

Risk Sentiment Secret Sauce?

Count SIX keywords:

- risk
- risks
- risky
- uncertain
- uncertainty
- uncertainties

Source: Feng Li 2006

Returns by Risk Sentiment Decile

Panel A – EW returns (128 months)

	Excess returns	
	Mean	t-statistics
Ret0	1.37	[2.36]
Ret1	1.52	[2.31]
Ret2	1.65	[2.71]
Ret3	1.26	[2.06]
Ret4	1.46	[2.33]
Ret5	1.32	[2.16]
Ret6	1.02	[1.66]
Ret7	1.38	[2.00]
Ret8	0.85	[1.34]
Ret9	0.78	[1.33]
Ret10	0.60	[0.95]
HGRET1	8.92	[3.35]
HGRET2	6.77	[3.69]

Source: Feng Li 2006

LS Risk Sentiment Portfolio Alpha

Figure 1 – value-weight hedge portfolio returns (HGRET) by calendar year

hedge portfolio return (%)

Year

Note: This graph plots the annual (compounded) returns of the value-weighted portfolio HGRET1, as defined in Table 3.

Source: Feng Li 2006

A Poster Child Example of News Disintermediation

(Full version available on "Nerds on Wall Street")

<http://nerdsonwallstreet.typepad.com/>

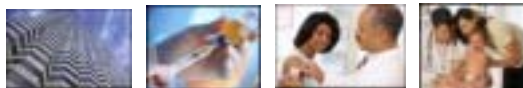
A microscope for examining how news affects markets.



News At Move Strong efficiency



Looking through the Microscope Oct 19, 2006. ABPI +57%



News and "Pre-News": Accentia +57%, Oct 19, 2006



Google finance, 4 pm Oct 19

A month earlier... ...in a specialized news source



<http://jncicancerspectrum.oxfordjournals.org/cgi/content/abstract/jnci.98/18/1292>

News, as good as it gets.

Specific immune response was associated with a dramatic and highly statistically significant increase in disease-free survival. This is the first formal demonstration of clinical benefit associated with the use of a human cancer vaccine.

Message posters found it. Oct 13



Where it started. 13 months prior.



Dis-intermediated News, 9AM Oct 19

Press Release Source: Accentia Biopharmaceuticals Inc.

New Landmark Study Shows the Use of a BiovaxID Formulation Improves Disease Free survival in 80% of Patients Whose Cancer has Relapsed after Chemotherapy Administered With or Without Rituxan

TAMPA, Fla.-(BUSINESS WIRE)-Accentia Biopharmaceuticals Inc. (NASDAQ: [ABPI](#)) - [News](#) reports that a new study by Dr. Susana Inoges and colleagues at the University of Navarra (Journal of the National Cancer Institute, 98: 1302-1301, Sept 20, 2006) has shown that the administration of a BiovaxID(TM) formulation given to patients with relapsed Follicular Non-Hodgkins Lymphoma following chemotherapy, with or without concomitant Rituxan(TM)(1), can induce complete long-lasting remissions. Moreover, these second remissions are uniquely characterized by a duration that exceeds the duration of the first remission (P less than .0001) in the study, 80% of patients achieved an immune response to the BiovaxID formulation, and among these responders, the median time of complete tumor remission has not been reached after 33 months of mean follow-up.



This took ~45 minutes to be reflected in price

ABPI, Oct 19, 2006

Mainstream Media #1: 2:24 pm



MSM #2, Nothing, at 4pm

REUTERS



MSM
 There are no recent news headlines for Accerta Biopharmaceuticals Inc. Please select any Developments or Press Releases for informative highlights on ACB.

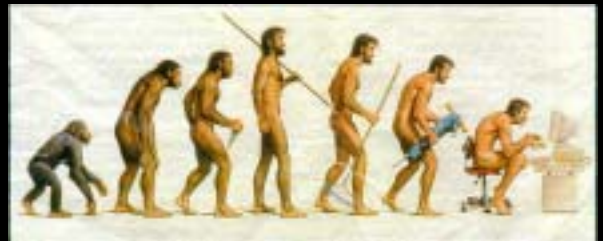
Clinical Trials Lessons

- There are thousands of ongoing trials, at hundreds of centers around the world
- The news dissemination process can be slow
 - And the news does not always mention the company
- Textbook example of the need for “metaknowledge”
 - Map trials to firms

Key Concepts

- Persistent Search, Disintermediation
- Deep Web – Database access (eg SEC)
- Semantic Web – XML
- “Molecular view” >> “Atomic view”
- ELINT
 - Traffic Analysis, NLP
- Overview, filter, alert, drill to details
 - HCI Challenges, Smoke/Fire

Significant progress evident



Persistent Search Using Yahoo Pipes



Source: R. Pasarella

Pipes Analysis of Pre-orders for THQ's Ratatouille Game Aug 07 – Deep Discounts in the UK, Weak early buzz

	Ratatouille Availability: Preorder - 26 Sep 2007 Publisher: Not Applicable	Rating: BBFC - TBC Platform: Wii Category: Other	RRP £29.99 Save £10.00 £19.99 Preorder
	Ratatouille Availability: Preorder - 26 Sep 2007 Publisher: THQ	Rating: BBFC - TBC Platform: Xbox 360 Category: Action	RRP £29.99 Save £10.00 £19.99 Preorder
	Ratatouille Availability: Preorder - 26 Sep 2007 Publisher: Not Applicable	Rating: BBFC - TBC Platform: PlayStation 2 Category: Other	RRP £29.99 Save £10.00 £19.99 Preorder
	Ratatouille Availability: Preorder - 26 Sep 2007 Publisher: Not Applicable	Rating: BBFC - TBC Platform: DS Category: Other	RRP £29.99 Save £10.00 £19.99 Preorder

Source: R. Pasarella



From Google Labs

Google Labs: Google's technology playground.

- Experimental Search** - Search and explore Google Search
- Google Code Search** - Search code across all public repositories
- Google Maps Trends** - See what's hot in Google Maps
- Google Trends** - See what's hot in Google
- Google Extensions for Firefox** - Google Browser Extensions
- Google Desktop** - Search your desktop

Alternate views for search results

Try out one of these options:

- View as a timeline
- View as a map
- View additional info

Keyword suggestions

Google Labs - Directed Search

Left-hand search navigation

Direct usage in a particular type of content - products, products or news, for example Google results across all results of content. This can be used by relevance.

Try out one of these options: **use best results**

RSS Integration

Be in N places at once, persistently.

Read before coding: NLP Science

Introduction to Information Retrieval

Free book here:
<http://www.csl.stanford.edu/~hinrich/information-retrieval-book.html>

**Read before trading:
Active Portfolio Management**



**Traditional
Semi-mandatory,
Semi-inspirational
Closing Quotation**

**Progress may have been all right
once, but it's gone on far too long.**

- Ogden Nash



Expect errors

Expect surprises



Expect strange innovations



Convair Flying Car, 1947



Expect success!

Center for Innovative Financial Technology

- UC Berkeley Haas School of Business
- DOB ~ Feb. 1, 2008
- People, $t=0$
 - David Leinweber
 - Terry Hendershott
 - John O'Brien
- Distinguished Advisors, $t=0$



Nerds on Wall Street
 A weekly blog for market and financial professionals, created by "Computer Scientist and ex-quant" David Leinweber, with contributions from elsewhere in the tech and VC world.

► About
 ► Archives
 ► February 2007

► Categories
 ► All categories
 View All 205 (24) Items

Powered by Typepad

FEBRUARY 17, 2007
Information-Driven Price Moves - A Case Study
 Download Casebook List

This is a follow-up to one of my previous posts, discussed and incorporated in financial markets in a way of great fun to some and slight "nerd" derision. The case involves a certain high-frequency trader, who had the largest percentage gain in US stock markets on the morning of Oct 19, 2006.

Several interesting (and related) facts:

- How should investors and traders, and their information providers, should deal with discontinuous information?
- There is more noise than people can handle, and for many (with some caveats) noise. Some more beyond basic forward events is needed.
- See: something like this one, which has language that would be of interest and "fun" stuff at the end for "nerd" and "ex-quant" people on search!

<http://nerdsonwallstreet.typepad.com/>

