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

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 64-75

A Short History of Market Information Technology
NYSE in 1792. The Buttonwood Tree

NYSE moves indoors. Tontine Coffee House. 1794

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

Beyond Ticker Tape
Exchange President Keith Funston with first NYSE computer.
1966

Disintermediation of Execution

Electronic Disintermediation of Execution

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)


Institutional Investor Alpha, February 2007

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

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
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 – Creatures of the net
  - Websites, mail, messages, chat, blogs, RSS...
  - "Social Media"

Investing, Trading & the Internet

Time isn’t what it used to be.


Earnings Surprises ...After WWW. (1995-1998)
The Text Frontier: *Finding Alpha on the Web*

Democratization and Disintermediation of Information

Ace Reporters?  
Who needs 'em!  
Do it Yourself  
Disintermediated News

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?
eAnalyst Case Study

News, Language Models, One day horizon

---

**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

---

**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

---

**Language Modeling**

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...

---

**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
More Than Words: Quantifying Language (in News) to Measure Firms’ Fundamentals
Tetlock, et al, UT Austin, Sep. 2006

News, Sentiment, One day horizon

News Sentiment (Tetlock)
- WSJ & DJNS Stories, S&P500, 1984-2004
- General Inquirer for sentiment
  - Academic system, psychology/linguistics
  - Developed over 20+ years
    - US NSF and Australian Research Council funds
    - Originally, PL/I on IBM mainframes
    - Try it at www.wjh.harvard.edu/~inquirer/
- PSTV and NGTV word scores
- One day LS trading simulation

Top of General Inquirer NGTV

Top of General Inquirer PSTV

Distribution of Annual Returns to Tetlock LS Trading Simulation (1984-2004)

Tetlock News Event Study (1984-2004)
Note Huge Pre-event Information Leakage!

Note: Before trading costs
Tetlock’s Results

Predictability & Consistency

The results in Table III are negative words in firm-specific news stories strongly predict lower returns on the following trading day. The coefficients on negative news words are strongly negatively significant at all 12 of the regressions where news stories from FTVS are included.


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

Deep web – Where’s the SEC?

Plain Old Ugly Edgar

SEC gets the XML Religion
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...

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

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. Diffing 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.

Lot’s of new discussion about derivatives.
An accounting tale of great complexity.

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

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

Returns by Risk Sentiment Decile

Source: Feng Li 2006

LS Risk Sentiment Portfolio Alpha

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

A month earlier…
…in a specialized news source

http://jnccancerpectrum.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

This took ~45 minutes to be reflected in price

Mainstream Media #1: 2:24 pm
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

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

Source: R. Pasarella

Source: R. Pasarella
From Google Labs

Google Labs – Directed Search

RSS Integration
Be in N places at once, persistently.

Read before coding:
NLP Science

Free book here:
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

http://nerdsonwallstreet.typepad.com/

Cover poll for forthcoming Wiley book

Questions?

You want proof? I’ll give you proof!

Lunch