Event Trading and the “New News”

David Leinweber* and Jacob Sisk**

*Leinweber & Co. and LBL, **Thomson Reuters

QWAFAFEW San Francisco, April 26, 2011

Running 47 years, since 1964.

Much closer to the holy grail of computers and language

Text Data is Superficial

An iceberg is a large piece of freshwater ice that has broken off from a snow-formed glacier or ice shelf and is floating in open water.

Dan Klein, UC Berkeley
... But Language is Complex

An iceberg is a large piece of freshwater ice that has broken off from a snow-formed glacier or ice shelf and is floating in open water.

Watson: Relationships and Lots of Data

More Data: Machine Translation

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td>That would be an interim solution which would make it possible to work towards a binding charter in the long term.</td>
</tr>
<tr>
<td>9x DATA</td>
<td>[this] [constituental] [assistance] [transitional] [who] [permitted] [licences] [to] [term] [to] [a] [charter] [to] [value] [contraignants]</td>
</tr>
<tr>
<td>10x DATA</td>
<td>[it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to] [a] [charter] [legally binding]</td>
</tr>
<tr>
<td>100x DATA</td>
<td>[this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding]</td>
</tr>
<tr>
<td>1000x DATA</td>
<td>[that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter]</td>
</tr>
</tbody>
</table>

Not just yet...

Predictions

<table>
<thead>
<tr>
<th>Today</th>
<th>2020 (likely)</th>
<th>2020 (hopefully)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find information</td>
<td>Synthesize information</td>
<td>Infer information</td>
</tr>
<tr>
<td>Keywords and names</td>
<td>Entities</td>
<td>Concepts</td>
</tr>
<tr>
<td>Knowledge-free “structural” systems</td>
<td>Knowledge from text</td>
<td>Knowledge from grounded contexts</td>
</tr>
<tr>
<td>“Talk” to search engines</td>
<td>Talk to embedded devices</td>
<td>Talk to mobile robots</td>
</tr>
<tr>
<td>Superficial patterns</td>
<td>Deep understanding</td>
<td>Monologs → dialogs</td>
</tr>
</tbody>
</table>
Back to 2011....

We know news moves markets.  
**An accidental experiment**

UAL September 7, 2008. Old news rises from the news crypt. 
(source, "Nerds on Wall Street", pp 206-208)

**Finding Alpha in News**  
*Background & Basics*

Chapter 4  
*Where Does Alpha Come From?*  
Life is math. The math is ideas.  
—Thomas Tsiolkovsky

Chapter 9  
The Text Frontier  
*AI, IA, and the New Research*  

*Nerds on Wall Street: Math, Machines, & Wired Markets*  
(May, 2009)  

**News and Stock Markets:**  
The Bad News & The Challenge

**Simple Positive/Negative Sentiment Classification on TRNA Tagged News**  
*Remarkably Similar Results*

**The Challenge:**  
*Move this line to the left to capture more alpha by filtering news*
A Global Pattern: Japan, Germany, Hong Kong, UK

Similar Patterns: France, Australia

Preview

- The old adage “buy on the rumor, sell on the news” is still true, much of the time, in many markets
  - Lots of “me, too” and “cat’s out of the bag” effects make unfiltered news hard to trade on profitably
- BUT, modern news systems include expanded news and “metadata” about news that allows you to set multidimensional hurdles for exploitable news.
- These hurdles, are a source of alpha for quants, input to trading algos, and a “Too Much Information” relief aid for all.

Significant improvements in electronic news

- More Sources, More Timely, More People Globally
- Metadata, for each story
  - Sentiment – positive or negative?
  - Relevance – about a specific firm?
  - Novelty – are there previous similar stories
  - Topic and types
- News Improvements
  - Volume of news
  - Breadth of news
    - how many firms covered
  - Depth of news
    - how much on each firm

Increasing news volume for US stocks

Increasing depth – news items per stock
Increasing breadth of coverage – more stocks

Evaluating news in trading

- Event studies
  - Set hurdles for news events
    - Volume, sentiment, novelty, relevance
  - Look at subsequent alpha (excess return) measured from event time
- Portfolio simulation
  - Same hurdles
  - But look at simulated trades (including transaction costs) in calendar time
- Both show exploitable alpha in news

"Low News Hurdle" event study

"High News Hurdle" event study

News Driven Portfolio Simulation

- Pure News signal driven
  - Long/Short, but not market neutral
  - Would be combined with many others: See Deutsche Bank report
- S&P 1500 Universe
  - Jan 2006 – Nov 2009
- Trading
  - At market close, on "high bar" sentiment signals
  - 25 bp transaction costs included
- Hold for 20 days, subject to
  - stop-loss rule set at 5%
  - profit take rule set at 20%

News Portfolio Simulation: Cumulative Return 2006-2009

News Improvements & Alpha Jump: Coincidence?

Independent Validation – Deutsche Bank 2010 (Pleasant surprise #1)

Remarkably similar return patterns, Independent Results!

Really sticking your neck out department.

• This work was completed at the end of 2009.
  - And we are all data miners if we look twice
• What happens if we use the same strategy in a true Out of Sample Test?
  - Generate signals on completely unseen news
  - Trade on completely unseen prices
• Drum roll please……

Where we left off

Pure News Analytic Portfolio: Simulated + True Out of Sample Results

Pure News Analytic Portfolio: True Out of Sample Results: 2010 YTD (Oct)
Pre e-Break Summary

- News sources have evolved dramatically.
  - Coverage, Analytics & Metadata, Speed
- Event studies show the promise of exploitable alpha
- Portfolio simulations confirm this, in calendar time
  - Using pure news signals, unclouded by overlays
  - Not just for quants.
  - Event frequencies and response time suitable for traditional managers
- Much more to talk about

Organization of this material

| Time Scale          | Human Intelligence Amplification | "Pure News Quant"
|---------------------|----------------------------------|------------------------
| Slower investment time scale. Multi-day | (Generating pure news signals - w/o mixing in other information) | "Pure News Quant" |
| Fast intraday       |                                  | "Pure News Quant" |

Event Study Research Automation with Visual Data Exploration

<table>
<thead>
<tr>
<th>Slow investment time scale</th>
<th>Fast intraday</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVENT STUDY EDA TOOLS</td>
<td></td>
</tr>
</tbody>
</table>

Event Study Explorer in Spotfire. Touch, Don't Just Look.

News Metric Design

A new manager skill set
Interactive Analysis:
Event Studies on Steroids

News Event Study Movies
Backup Plan

ESE Movie 1 - Overview
ESE Movie 2 – Filter Control Panel

ESE Movie 3 – HF Rolling Window
ESE Movie 4 – Topic Codes
Turning the knobs on analytic news filters

- News based
  - Volume of news
  - Novelty
  - Relevance/Sentiment
  - Metadata
    - Type, topic
- Set of stocks
  - Sector
  - Capitalization
- Expectation: More stringent filters will produce fewer events, but with higher returns

Interesting (and expected) EDA Result:
News is more exploitable for small stocks

Sample Intraday Event Study Results

- Event Definitions:
  - LONG EVENT if relevance*sent_pos>.5 and
  - SHORT EVENT if relevance*sent_neg>.5
- Details:
  - Period: 2009-01-01 - 2010-10-01
  - Data: TRNA v2
  - Universe: S&P1500
  - Time horizon: 180 minutes

Intraday Event Studies

Still on a human sensible time scale.
Minutes to hours, not milliseconds.

Robust signals over time, except for major meltdown days.
Turning the Time Honored Sector Knob

FINANCIALS

FINANCIALS (Re Scaled)

BASIC MATERIALS

ENERGY

Turning the news genre knob
Item_genre: BRIEF

**An EDA Discovery!**

- More updates = More alpha
- Why?
  - More *important news* is updated more times
  - A story about a new spark plug get no updates
- Importance is not a current metadata item
  - Seems valuable, other ways to measure

---

**Item_genre: RPT**
Item_genre: UPDATE 6

How about replacing these two UPDATE=0 and >0 Basis Points?

Item_genre: UPDATE 8

How about replacing these two UPDATE=0 and >0 Basis Points?

Item_genre: UPDATE 3 (rescaled)

How about replacing these two UPDATE=0 and >0 Basis Points?

Item_genre: UPDATE 4 (rescaled)

How about replacing these two UPDATE=0 and >0 Basis Points?

Item_genre: UPDATE 6 (rescaled)

How about replacing these two UPDATE=0 and >0 Basis Points?

Item_genre: UPDATE 8 (rescaled)

How about replacing these two UPDATE=0 and >0 Basis Points?
Turning the topic code knob

What news event studies tell us about investment and trading

- Previous multiday/week results
  - Many slow >100+ bp signals
  - Ample time to accumulate positions
  - Smaller caps stronger. Attention hypothesis
- Many short term ~50 bp responses in < 3 hrs
  - Opportunistic trading
  - Algo refinement

<table>
<thead>
<tr>
<th>SENTIMENT INDICES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Intelligence Amplification</td>
</tr>
<tr>
<td>Slow investment time scale</td>
</tr>
<tr>
<td>Fast intraday</td>
</tr>
</tbody>
</table>
**Sentiment Indices:**

*News Analytics Beyond Quant Equity*

- **Asset allocation** decision support
  - Large macro scale decisions. High Value
  - Country Allocations
  - Global Sector Allocations

- Sentiment can be aggregated using any Indices
  - **Or custom portfolios of your holdings**

**Aggregate Sentiment Indices:**

*Built From Individual Stock Metrics*

- Metrics are calculated for a single stock
- Taking metrics for all of the stocks in an existing index (like the SP500, TOPIX, or a TR sector), and summing, averaging

\[
\sum_{i=1}^{n} w_i e_i
\]

**Sentiment “Big Picture” Heatmaps Using Actual Historical TRNA/RNSE Data**

- 8 countries
  - US, CAN, UK, FR, GER, JPN, HK, AUS
- 10 Reuters sectors
- Period: Jan 1, 2006 – Jan 31, 2010

- **Aug 21, 2008**
  - US concerns, yet to spread globally

- **Dec 10, 2008**
  - Global markets follow
March 13, 2009
Fear abating, 2009 Q1

Sentiment Index Heatmap
Friday, March 13, 2009

Dec 14, 2009
Growing sense of relief 2009 Q4

Sentiment Index Heatmap
Monday, December 14, 2009

Drill Down To Details
Top Positive/Negative news events, by stock

Intelligence Amplification:
Anti-“Too Much Information” Weapon
For All Types of Managers

- Intelligent agent to assist with TMI
- See big trends, drill down to details
- Supports the decisions managers make, the way they make them
- Allow application of manager skill

PURE NEWS
PORTFOLIO SIMULATIONS

<table>
<thead>
<tr>
<th>Human Intelligence Amplification</th>
<th>“Pure News Quant”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow investment time scale</td>
<td>PORTFOLIO SIMS</td>
</tr>
<tr>
<td>Fast intraday</td>
<td></td>
</tr>
</tbody>
</table>

Back to where we started, once more, with feeling…

News Improvements & Alpha Jump: For us and for D-Bank. Coincidence?

Alpha from advanced news technology
Remarkably similar return patterns, Independent Results! Deutsche Bank

Figure 53: Cumulative performance (after cost) for simulated news sentiment (bars)

Where we left off

Pure News Analytic Portfolio: Simulated + True Out of Sample Results

Event Study Explorer Free Test Drive - BYOD

Event Study Explorer

Daring Live BYOD Demonstrations

EDA Demonstrations

- Extreme news day event study explorer
- High(er) frequency intraday event studies
- Visual sentiment metric construction for
  - individual securities
  - sectors and markets
Back to the future....

What would be in Watson's portfolio in 2020?

Predictions

<table>
<thead>
<tr>
<th></th>
<th>Today</th>
<th>2020 (likely)</th>
<th>2020 (hopefully)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find information</td>
<td>Synthesize information</td>
<td>Infer information</td>
<td></td>
</tr>
<tr>
<td>Keywords and names</td>
<td>Entities</td>
<td>Concepts</td>
<td></td>
</tr>
</tbody>
</table>
| Knowledge-
  free
  "structural" systems | Knowledge from text | Knowledge from grounded contexts |
| “Talk” to search engines | Talk to embedded devices | Talk to mobile robots |
| Superficial patterns | Deep understanding | Monologs → dialogs |

Loosely
Structured Information (text and relationships)

Highly
Structured data (HF ticks)

Many connections


Mauricio A. Hernandez, Howard Ho, Georgia Koukiou, Rajasimha Krithivasan; Lucas Pons, Irene R. Santos; Wshikumar Vashishth, Sayor Dar

Abstract

We present Mida, a system that uses complex data patterns to reason about aggregated financial risk features. Based on data from a large number of financial institutions and regulated firms, we present Mida's ability to perform this analysis in one of the largest and most comprehensive risk management systems in the world. Mida's ability to perform this analysis in one of the largest and most comprehensive risk management systems in the world.

Interbank Lending
Extracted from FDIC & SEC

(IBM MIDAS 2010)
Interbank Lending Extracted from FDIC & SEC

IBM MIDAS 2010

Midas representation of co-lending network in 2005. "We see that there are three large components of co-lending, and these hub banks, with connections to the large components. There are also satellite hubs. In order to determine which banks in the network are most likely to contribute to systemic failure, we compute the normalized eigenvalue centrality score" (Hernando et al., 2010)

"Comparing 2006 with 2005 (Exhibit 2.1), we see that there were disjointed large components connected by a few central nodes. From 2007 onwards, as the financial crisis begins to take hold, co-lending activity diminished markedly. Also, all high centrality banks tend to cluster into a single large joint component in the latter years."

Lost of things computers can’t do is getting shorter, fast.
Maybe Kurzweil is right about all those vitamins.

“Nerds on Wall Street: Math, Machines, & Wired Markets” (Wiley, 2009)
Amazon: [Link](http://www.amazon.com/Nerds-Wall-Street-Math-Machines/dp/0470179068)