

Which News Moves Stock Prices? A Textual Analysis

Jacob Boudoukh – IDC

Ronen Feldman – Hebrew University

Shimon Kogan – University of Texas & IDC

Matthew Richardson – NYU & NBER

October, 2013 – Q Group Fall Seminar

Motivation

- Basic tenet of rational asset pricing:
returns are a function of new information
- Supporting evidence came mainly from event studies
 - Ball Brown (1968) on earnings
 - Fama Fisher Jensen Roll (1969) on splits
 - ...
- Roll's (1988) presidential address showed a weak link between news and prices, along with
 - Shiller (1981)
 - Cutler, Poterba, and Summers (1989)
 - ...

Main Results

- Roll posits that if information moves prices
 $R^2(AllDays) \ll R^2(NoNewsDays) \approx 100\%$
however, he finds that
 $R^2(AllDays) = R^2(NoNewsDays) = 20\%$,
while
- Our story: the long-standing puzzle is a result of using poor proxies for true relevant news
- We exploit recent innovations in textual analysis to show that
 - when able to identify true relevant news and their importance, news matter!
 - different forms of news give rise to different patterns of continuations and reversals

Standard Approach

Standard approach towards identification of publicly released news (e.g., Tetlock (2007)) is to

- Obtain text data (e.g., news articles)
- Link text with firms through name/ticker mention
- Identify the tone of the text by signing each word with a predetermined dictionary (e.g., positive vs. negative) and summing up signs

The Stock Sonar (TSS)

- A proprietary information extraction platform specific to finance, available freely on the web and through Dow Jones
- Differs from existing methodologies in the finance literature:
 - 1 Uses a dictionary but refines IV-4/Loughran&McDonald (2011) to include sentiment modifiers, e.g., “highly” versus “mostly”
 - 2 Operates at the phrase level, not just on word, e.g., double negatives (e.g., “reducing losses”), connectors (e.g., “despite”)
 - 3 Sorts through the document and parses out the meaning in the context of possible events relevant to companies, and determines which companies map to the events.

TSS Example



Event List

Acquisition

Merger
Tender_Offer

Analyst Recommendations

Analyst_Expectation
Analyst_Opinion
Analyst_Rating
Analyst_Recommendation
Credit_OR_Debt_Rating
Fundamental_Analysis
Price_Target
Rating_Agency_List
Technical_Analysis

Deals

Contract_OR_Agreements
Exclusivity
Licensing
MOU
Negotiation
Pact
Service_OR_Product

Employment

Board
Chair
Executive
Senior_Executive
Workforce

Financial

Dividend
Stock_Financial
Reports
Forecasts
Metric_Change
Investment
Derivatives
Financing

Legal

Bankruptcy
Investigation
Judgement
Lawsuit

Partnerships

Alliance
Joint_Venture
Termination

Product

Approval
Discontinuation
Expansion_of_Line
Issues
New_Product
Recall
Submission
Testing_OR_Trial
Updates_OR_Upgrades

Data Description

- Dow Jones Newswire articles (with at least 50 words)
- S&P500 firms (as of beginning of each year), 2000-2009
- For each article, we obtain
 - 1 Ticker(s)
 - 2 Event(s)
 - 3 Tone on a scale of $(-1, +1)$
 - 4 Time stamp
- Restructure the data to follow <ticker-date> format, using a 1530 cutoff time

Summary Stats

	# Obs.	# Obs. with event count			# Tickers	# Articles (daily)	# Words (per art.)	# Relv. Words (per art.)
		=1	=2	>3				
Total	1,229,359	124,158	26,600	7,422	791	3.6	325	59
No News	696,985	NA	NA	NA	790	NA	NA	NA
Unid News	374,194	NA	NA	NA	791	2.6	329	49
Idea News	158,180	124,158	26,600	7,422	790	6.1	316	81
Acquisition	22,270	11,811	7,068	3,391	724	8.6	302	76
Analyst Rec	12,411	6,665	3,920	1,826	680	8.5	335	66
Deals	30,101	17,001	8,790	4,310	718	6.8	315	93
Employment	21,489	14,024	5,269	2,196	741	6.3	283	87
Financial	69,205	48,873	14,989	5,343	783	7.6	309	71
Legal	10,764	6,244	2,881	1,639	581	8.6	291	71
Partnerships	10,047	4,765	3,347	1,935	587	7.3	371	110
Product	25,181	14,775	6,936	3,470	652	7.1	366	108

Unid News: news coverage days without *any* corporate events

Idea News: news coverage days with *a* corporate event

Event Frequency

Return rank	<i>p0-p10</i>	<i>p10-p30</i>	<i>p30-p70</i>	<i>p70-p90</i>	<i>p90-p100</i>
# of articles	4.5	3.5	3.3	3.5	4.4
# of words	1,445	1,140	1,093	1,133	1,423
% of rel. words	17.5%	16.6%	16.5%	16.6%	17.3%
No News	-6.6%	0.5%	2.8%	0.5%	-6.5%
Unid News	2.2%	-0.4%	-0.4%	-0.4%	1.1%
Iden News	30.8%	-5.7%	-10.7%	-5.4%	34.2%
Acquisition	18.8%	-4.2%	-6.2%	-5.4%	25.3%
Analyst Rec	71.1%	-11.9%	-22.1%	-10.3%	61.9%
Deals	6.5%	-4.0%	-3.1%	-1.9%	17.6%
Employment	20.5%	-2.1%	-5.8%	-2.8%	12.3%
Financial	66.2%	-11.1%	-23.0%	-10.3%	68.2%
Legal	16.8%	0.2%	-5.1%	-4.4%	11.9%
Partnerships	3.3%	-2.3%	-1.0%	-2.9%	11.2%
Product	10.4%	-2.2%	-4.1%	-2.6%	15.6%

1

¹Stock and year assignments

Variance Ratios

	<i>p</i> 10	<i>p</i> 25	Median	<i>p</i> 75	<i>p</i> 90	<i>N</i>
Unid News	0.75	0.95	1.2	1.5	2.0	754
Iden News	1.05	1.49	2.2	3.3	5.1	672
Acquisition	0.64	1.01	1.5	2.7	6.3	294
Analyst Rec	1.05	1.70	2.7	5.1	13.5	190
Deals	0.65	0.94	1.4	2.1	4.0	273
Employment	0.65	0.96	1.4	2.4	4.1	329
Financial	1.29	1.95	2.9	4.7	7.6	581
Legal	0.46	0.71	1.3	2.3	4.9	116
Partnerships	0.49	0.69	1.1	1.5	2.0	110
Product	0.62	0.86	1.3	2.1	3.1	223

Variance Ratios show more formally that when we identify news, news matter for stock returns

Roll R^2

- We go back to Roll's (1988) analysis: do factor models' R^2 change across days with and without news?
- He posits:
 - returns on no-news days should be dominated by systematic risk factors
 - ... while returns on news days should not be
- Roll finds R^2 s that are similar on both types of days!
- We follow a similar procedure using both 1 and 4 factor models (including size, value, and momentum factors)

Roll R^2 Results

	Firm Level			Industry Level		
	Mean R^2	Med R^2	N	Mean R^2	Med R^2	N
1 Factor						
All	28.6%	27.8%	791	28.5%	27.6%	60
No News	32.1%	33.3%	774	32.3%	34.1%	60
Unid News	30.8%	30.3%	721	30.9%	30.1%	58
Iden News	18.5%	15.9%	597	17.3%	16.2%	55
$\frac{NoNews}{IdenNews}$	1.73	2.10		1.87	2.10	
4 Factors						
All	33.6%	32.7%	791	31.8%	31.0%	60
No News	38.0%	38.6%	774	36.3%	36.9%	60
Unid News	35.8%	35.9%	721	34.0%	33.6%	58
Iden News	22.3%	19.6%	597	19.2%	17.1%	55
$\frac{NoNews}{IdenNews}$	1.71	1.96		1.89	2.16	

For Unid News days, Roll's results hold, but not for Iden News days!

Daily Tone

	Event count	N	mean	TSS				N	mean	IV4				IV4-TSS
				p10	p50	p90	p90-p10			p10	p50	p90	p90-p10	Corr.
Unid News	374,194	135,643	0.25	-0.50	0.50	0.75	1.25	364,969	0.33	0.01	0.32	0.69	0.68	0.33
Iden News	158,180	158,097	0.38	-0.40	0.50	0.83	1.23	158,150	0.32	0.08	0.31	0.58	0.50	0.34
Acquisition	22,270	22,228	0.54	0.25	0.50	0.80	0.55	22,238	0.36	0.06	0.36	0.67	0.61	0.19
Analyst Rec	12,411	12,371	0.06	-0.67	0.00	0.75	1.42	12,401	0.23	-0.06	0.22	0.54	0.60	0.32
Deals	30,101	30,082	0.60	0.50	0.67	0.83	0.33	30,084	0.42	0.12	0.43	0.69	0.57	0.19
Employment	21,489	21,482	0.42	-0.41	0.50	0.80	1.21	21,475	0.30	-0.07	0.33	0.64	0.71	0.17
Financial	69,205	69,124	0.29	-0.54	0.50	0.83	1.37	69,156	0.29	-0.01	0.29	0.62	0.63	0.38
Legal	10,764	10,761	-0.21	-0.80	-0.40	0.57	1.37	10,763	0.16	-0.16	0.17	0.50	0.66	0.31
Partnerships	10,047	10,040	0.63	0.50	0.67	0.83	0.33	10,040	0.44	0.16	0.46	0.69	0.53	0.24
Product	25,181	25,166	0.55	0.18	0.67	0.83	0.65	25,155	0.34	0.00	0.37	0.65	0.65	0.31

Comparing TSS with IV4:

- Correlation is positive but far from 1
- TSS mean vary more across event types
- TSS spread is larger within an event type

Roll R^2 with Tone

	Factor(s) Only Mean R^2	with TSS Scores Mean R^2	with IV4 Scores Mean R^2	N	ΔR^2 (TSS)
1 Factor					
Unid News	30.9%	30.9%	31.0%	58	1.00
Iden News	17.3%	18.9%	17.5%	55	1.09
New News	15.7%	17.9%	16.0%	55	1.14
Acquisition	13.1%	14.7%	14.3%	31	1.12
Analyst Rec	17.3%	25.3%	21.5%	32	1.46
Deals	20.6%	25.7%	26.3%	34	1.25
Employment	19.4%	24.7%	23.5%	37	1.27
Financial	14.1%	17.0%	15.0%	52	1.20
Legal	16.9%	23.5%	20.8%	23	1.39
Partnerships	25.0%	29.7%	28.4%	17	1.19
Product	28.2%	34.3%	34.0%	30	1.22

Including tone improves model's R^2 by 12% – 46%.

Return Predictability

- Evidence so far supportive of market efficiency
- Behavioral finance predicts:
 - “real” news $\Rightarrow$ continuation (i.e., partial adjustment)
 - “fluff” news $\Rightarrow$ reversal (i.e., over reaction)
- Big literature on this:
Stickel and Verrachia (1994), Daniel, Hirshleifer and Subrahmanyam (1998), Hirshleifer (2000), Pritamani and Singal (2001), Chan (2003), Vega (2006), Barber and Odean (2008), Gutierrez and Kelley (2008), Tetlock, Sarr-Tsechansky, and MacsKassy (2008), Tetlock (2010, 2011)...
- Our ability to differentiate real vs. fluff, offers a special opportunity

Return-Based Predictability

- Each day, sign stocks based on day t return
 - large positive move ($> + \text{s.d.}$)
 - large negative move ($< - \text{s.d.}$)
- Strategy consists of
 - long \$1 across all large positive move stocks
 - short \$1 across all large negative move stocks
- Hold through the end of the following day

Reversals and Continuations

	Return Sorts			Return and Score Sorts	
	No News	Unid News	Idea News	Long Score	Short Score
Alpha	-0.161 [0.027]***	-0.062 [0.032]*	0.197 [0.041]***	0.342 [0.087]***	0.083 [0.082]
Mkt-rf	-0.197 [0.035]***	-0.164 [0.040]***	-0.203 [0.064]***	-0.369 [0.163]**	-0.048 [0.135]
SMB	0.067 [0.070]	0.132 [0.081]	0.057 [0.106]	0.266 [0.230]	0.143 [0.197]
HML	0.152 [0.075]**	0.06 [0.097]	-0.006 [0.111]	-0.458 [0.253]*	0.21 [0.225]
UMD	-0.064 [0.049]	-0.072 [0.060]	-0.012 [0.064]	0.156 [0.144]	-0.171 [0.114]
Observations	2487	2417	2030	751	740
R^2	0.046	0.021	0.022	0.106	0.015

Daily Statistics					
Mean	0.156%	0.059%	0.196%	0.322%	0.106%
s.d.	1.340%	1.545%	1.857%	2.522%	2.270%
SR	1.837	0.604	1.666	2.022	0.738
N Longs	35.8	21.7	11.2	7.8	7.8
N Shorts	38.0	20.6	10.5	7.7	7.7

Reversal

Continuation

We observe reversals following No News days and **continuations** following Idea News days

Reversals and Continuations

	Return Sorts			Return and Score Sorts	
	No News	Unid News	Iden News	Long Score	Short Score
Alpha	-0.161 [0.027]***	-0.062 [0.032]*	0.197 [0.041]***	0.342 [0.087]***	0.083 [0.082]
Mkt-rf	-0.197 [0.035]***	-0.164 [0.040]***	-0.203 [0.064]***	-0.359 [0.163]**	-0.048 [0.135]
SMB	0.067 [0.070]	0.132 [0.081]	0.057 [0.106]	0.216 [0.210]	0.143 [0.197]
HML	0.152 [0.075]**	0.06 [0.097]	-0.006 [0.111]	-0.458 [0.253]*	0.1 [0.225]
UMD	-0.064 [0.049]	-0.072 [0.060]	-0.012 [0.064]	0.116 [0.114]	-0.071 [0.114]
Observations	2487	2417	2030	75	70
R^2	0.046	0.021			

	Daily Statistics			Iden News Cont + Score	
	No News	Unid News	Iden News	Long Score	Short Score
Mean	0.156%	0.059%	0.196%	0.322%	0.106%
s.d.	1.340%	1.545%	1.857%	2.522%	2.270%
SR	1.837	0.604	1.666	2.022	0.738
N Longs	35.8	21.7	11.2	7.8	7.8
N Shorts	38.0	20.6	10.5	7.7	7.7

Focusing on Iden News days only, we document a strong tone-based return predictability

Complexity

Look for days/stocks with complex information structure

- more than 2 Iden articles (perhaps with similar events)
- more than 2 Iden events (perhaps in one or two articles)
- sentiment score dispersion (same event but different tone)

Complexity and Volatility

	Return Rank			Stock SD and Variance		
	20% Extreme	40% Moderate	40% Low	Med SD	N Tickers	Var Ratio
Total	1.0%	-0.6%	0.1%	2.64	791	1.16
No News	-6.6%	0.5%	2.8%	2.38	781	1.00
Unid News	1.6%	-0.4%	-0.4%	2.66	764	1.20
Idea News	32.5%	-5.5%	-10.7%	3.51	681	2.15
Comp	73.4%	-13.0%	-23.7%	4.49	470	3.67
Article Comp	82.2%	-14.3%	-26.8%	4.67	368	3.98
Event Comp	78.3%	-14.1%	-25.0%	4.53	401	3.72
Disag Comp	89.2%	-16.3%	-28.3%	4.55	391	3.69

Complex days are associated with **much** higher volatility

Complexity and R^2 s

	Returns – Firm Level			Returns – Industry Level			Score&Returns – Industry Level		
	Median R^2	Ratio (No News)	N	Median R^2	Ratio (No News)	N	Median R^2	Ratio (No Score)	N
Single Factor Regressions									
Total	27.8%	1.19	791	27.5%	1.20	60	27.9%	1.01	60
No News	33.3%	1.00	774	33.0%	1.00	60	33.0%	1.00	60
Unid News	30.3%	1.10	721	29.3%	1.13	58	29.3%	1.00	58
Iden News	15.9%	2.10	597	16.0%	2.07	55	18.0%	1.12	55
Comp	11.2%	2.97	335	12.2%	2.70	49	15.1%	1.24	49
Article Comp	10.8%	3.08	245	11.3%	2.92	41	14.6%	1.29	41
Event Comp	10.2%	3.25	259	9.1%	3.65	43	13.1%	1.44	43
Disag Comp	10.7%	3.10	264	7.5%	4.38	44	13.1%	1.74	44

R^2 spread is even more pronounced ($\times 3$ instead of $\times 2$)

Trading on Complexity

	Score and Complexity Sorts				
	All Idn	All Complex	Article Comp	Event Comp	Disag Comp
Alpha	0.068 [0.014]***	0.18 [0.039]***	0.191 [0.043]***	0.21 [0.043]***	0.231 [0.048]***
Mkt-rf	-0.095 [0.019]***	-0.191 [0.051]***	-0.164 [0.059]***	-0.212 [0.072]***	-0.249 [0.082]***
SMB	-0.001 [0.032]	0.069 [0.094]	0.027 [0.102]	0.02 [0.107]	0.123 [0.129]
HML	-0.254 [0.036]***	-0.394 [0.083]***	-0.449 [0.096]***	-0.461 [0.100]***	-0.454 [0.115]***
UMD	0.138 [0.022]***	0.254 [0.051]***	0.327 [0.059]***	0.225 [0.058]***	0.242 [0.059]***
Observations	2511	2198	1913	1919	1685
R^2	0.181	0.097	0.114	0.101	0.103

Daily Statistics					
Mean	0.060%	0.169%	0.185%	0.192%	0.222%
s.d.	0.785%	1.905%	1.976%	1.940%	2.061%
SR	1.206	1.404	1.481	1.564	1.705
N Longs	46.0	10.0	7.6	8.2	7.8
N Shorts	46.5	10.6	9.1	8.4	8.1

Complexity-based strategies yield substantial before-costs alphas ($\approx 45\%$ per year)

Conclusions

- When you can identify news and how important it is, news matters!
 - Results are stronger for complex days
- Events identification allows us to better understand when stock prices
 - over-react – no / “fluff” information
 - under-react – “real” / complex information