Information Processing in Financial Markets: Textual Analysis and Information Characteristics

UBS Quantitative Investment Conference 2013

Nicky Ferguson
University of Cambridge

nicky.ferguson@cantab.net
LinkedIn.com/in/nickyferguson

UNIVERSITY OF CAMBRIDGE
Judge Business School

UBS Disclaimer

This presentation was produced solely by Nicky Ferguson. The opinions expressed herein do not necessarily reflect the views and opinions of UBS. UBS accepts no responsibility for the accuracy, reliability or completeness of the information and will not be liable either directly or indirectly for any loss or damage arising out of the use of this presentation or part thereof.

Reproduced with permission.

Information Utilization – Why Study It?

"The most valuable commodity I know of is information, wouldn't you agree?"

- Efficient firm valuations should be equal to investors discounted information sets.
- Information sets contain quantitative information regarding fundamentals.
- However excess noise in stock returns cannot be explained by changes in firm fundamentals.
- Qualitative information about the firm is included in investors information sets, but this is difficult to measure!
- Investors are subject to limited attention -> information characteristics are important. News media play a key role in information dissemination.
- Elton (1999, JF) Stock price process with information shocks. $\mathbf{R_t = E[R_{t-1}] + I_t + e_t}$

Measuring Information – Academic Literature

- **Tetlock (2007, JF)** – examines the effect of tone at the market level, finds media pessimism predicts low returns.

- **Tetlock, Saar-Tsechansky, and Macskassy (2008, JF)** – examines the effect of tone at the firm level on stock returns and fundamentals. Finds negative words can predicts low earnings and stock returns.

- **Barber and Odean (2008, RFS)** – find that individuals are net purchasers of attention grabbing stocks which can temporarily inflate a stocks price.

- **Fang and Peress (2009, JF)** – Find stocks more heavily covered by the news media have lower returns in support of the informational risk hypothesis.

Measuring Information – Academic Literature

- **Loughran and McDonald (2011, JF)** – create a financial news specific psychosocial dictionary from 10-Ks with stronger predictive ability on stock returns than Harvard psychosocial dictionary.

- **Engelberg and Parsons (2011, JF)** – Causal impact of news media in financial markets using weather. **Peress (2012, INSEAD)** using newspaper strikes.

- **Griffin, Hirschey, and Kelly (2011, RFS)** – Global study, stronger stock price reaction to news in developed countries.

- **Boudoukh, Feldman, Kogan, and Richardson (2012, NBER)** – uses a dictionary and phrase level analysis and explains greater variation in stock price due to news.

- **Garcia (forthcoming, JF)** - predictability of stock returns by news media content is greatest during recessions. Using two columns from NYT on DJIA.

Measuring Information – Practicalities

- News articles in machine readable format
- Tagged data to allow matching and implementation
- Data sources:-
  - LexisNexis
  - Factiva
  - S&P Capital IQ Key Developments
  - Thomson Reuters – News Feed Direct
  - Dow Jones – Elementized News Feed (XML)
  - Bloomberg – Event Driven Trading Feed
- Text analysis – by dictionary, break news article into vector of words, and compare to dictionary, calculate measure of semantics
- e.g. R - comparing vectors of words - functions `%in%` and `which`

Measuring Information

Negative – 2337 words

ANTICOMPETITIVE
ANTITRUST
ARGUE
ARGUED
ARGUING
ARGUMENT
ARGUMENTATIVE
ARGUMENTS
ARREARAGE
ARREARAGES
ARREARS
ARREST
ARRESTED
ARRESTS
ARTIFICIALLY
ASSAULT
ASSAULTED
ASSAULTING
ASSAULTS
ASSERTIONS
ATTRITION
AVERSELY
BACKDATING
BAD
BAIL
BAILOUT
BALK
BALKED
BANKRUPT
BANKRUPTCIES
BANKRUPTCY
BANKRUPTED
BANKRUPTING
BANKRUPTS
BANS
BARRED
BARRIER
BARRIERS
BOTTLENECK
BOTTLENECKS
BOYCOTT
BOYCOTTED
BOYCOTTING
BOYCOTTS
BREACH
BREACHED
BREACHES
BREACHING
BREAK
BREAKAGE
BREAKAGES
BREAKDOWN
BREAKDOWNS
BRIBE

Positive – 353 words

ACHIEVEMENT
ACHIEVEMENTS
ACHIEVES
ACHIEVING
ADEQUATELY
ADVANCEMENT
ADVANCEMENTS
ADVANCES
ADVANCING
ADVANTAGE
ADVANTAGED
ADVANTAGEOUS
ADVANTAGEOUSLY
ADVANTAGES
ALLIANCE
ALLIANCES
ASSURE
ASSURED
ASSURES
ASSURING
ATTAIN
ATTAINED
ATTAINING
ATTAINMENT
ATTAINMENTS
ATTAINS
ATTRACTIVE
ATTRACTIVENESS
BEAUTIFUL
BEAUTIFULLY
BENEFICIAL
BENEFICIALLY
BENEFIT
BENEFITED
BENEFITING
BENEFITTED
BENEFITTING
BETTER
BOLSTERED
BOLSTERING
BOLSTERS
BOOM
BOOMING
BOOST
BOOSTED
BREAKTHROUGH
BREAKTHROUGHS
BRILLIANT
CHARITABLE
COLLABORATE
COLLABORATED
COLLABORATES
COLLABORATING
COLLABORATION

http://www3.nd.edu/~mcdonald/Word_Lists.html

Paper I

Ferguson, Guo, Lam, and Phillips (2012) Media Content and Stock Returns: The Predictive Power of Press. *Under Review – European Financial Management*

Measuring Information – Sample

- Sample collected from LexisNexis
- Firm specific articles classified using LexisNexis relevance rating (95%)
- Articles downloaded in XML format
- Count Positive and Negative words using Loughran and MacDonald lists
- Calculate measures:

$$Positive\ Content = \frac{number\ of\ positive\ words}{Total\ words}$$

$$Negative\ Content = \frac{number\ of\ negative\ words}{Total\ words}$$

Measuring Information - Timeline

- FTSE 100 firm specific news articles.
- Sample collected UK national newspapers 1981 – 2010.
- Newspapers are published and delivered before 7am each trading day.
- FT goes to press around 1am each trading day.

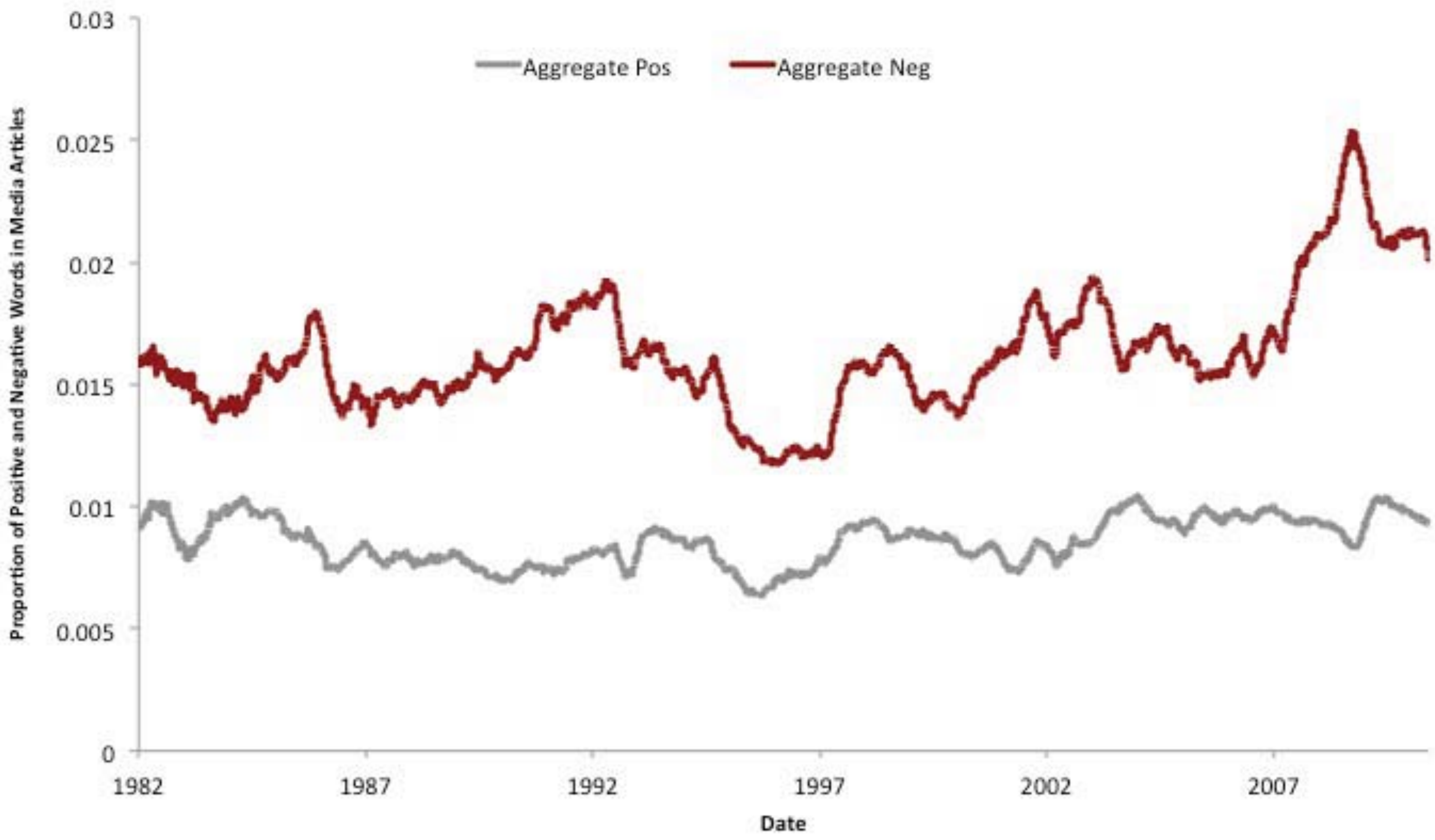
Data – Summary Statistics

Panel A: Sample statistics for raw media data

Year	Total Articles	Coverage				Average Article Words	Mean	Mean
		FT	*Times*	*Guardian*	*Mirror*		*Positive*	*Negative*
1981–1985	15431	82%	6%	12%	0%	442	0.0098	0.0165
1986–1990	30842	47%	20%	33%	0%	435	0.0078	0.0159
1991–1995	39284	51%	20%	27%	2%	548	0.008	0.0163
1996–2000	55596	51%	22%	16%	11%	476	0.0088	0.0154
2001-2005	40391	45%	19%	24%	12%	441	0.0091	0.0185
2006-2010	83103	66%	10%	18%	6%	476	0.0090	0.0222
1981–2010	264647	56%	17%	21%	6%	475	0.0087	0.0183

Time Variation In News Variables

- 100 day rolling average of news variables



Can We Predict Stock Returns?

- Company specific news articles published on day t are matched company specific close to close stock returns on day t. Results are displayed for log returns, and FFCAR(-252,-31)
- Measures of positive and negative news content are standardized.

	$Return_{+1,+1}$		$FFCAR_{+1,+1}$	
	FT	*ALL*	*FT*	*ALL*
Pos	0.1474***	0.0926***	0.1574***	0.0497***
	(5.51)	(6.98)	(6.32)	(4.11)
Neg	-0.0796***	-0.0551***	-0.0923***	-0.0235***
	(-5.37)	(-6.75)	(-6.65)	(-3.39)

- Control for past abnormal stock returns, size, b/m, turnover.
- Standard errors clustered by day.

- 1 s.d. increase in positive news content results in a 5 bps increase in abnormal returns.
- 1 s.d. increase in negative news content results in a 2.4 bps decrease in abnormal returns.
- Effect of positive news is statistically greater than negative news (χ^2, p=0.05)

Highly Visible News and Fundamentals

- Dependent variable is FFCAR(+1,+1)

	FT			*ALL*		
Pos	0.1483***	0.1285***	0.1302***	0.0392***	0.0331***	0.0239*
	(5.48)	(4.76)	(4.48)	(2.94)	(2.61)	(1.71)
Neg	-0.0955***	-0.0717***	-0.0771***	-0.0230***	-0.0204***	-0.0219**
	(-6.68)	(-4.86)	(-4.89)	(-2.92)	(-2.71)	(-2.57)
Fund	0.0001		0.0004	-0.0003		-0.0002
	(0.08)		(0.23)	(-0.42)		(-0.29)
*Pos*Fund*	0.0812		-0.0103	0.1313**		0.1134**
	(0.80)		(-0.09)	(2.57)		(2.19)
*Neg*Fund*	-0.0215		0.0047	-0.0527*		-0.0512*
	(-0.35)		(0.07)	(-1.93)		(-1.86)
MC		0.0015	0.0015		-0.0007	-0.0009*
		(1.13)	(1.06)		(-1.31)	(-1.66)
*Pos*MC*		0.1750*	0.1762*		0.1409***	0.1429***
		(1.77)	(1.68)		(3.52)	(3.36)
*Neg*MC*		-0.1727***	-0.1767***		-0.0191	-0.0067
		(-3.48)	(-3.43)		(-0.99)	(-0.33)

Firm Size and Valuations

- Dependent variable is FFCAR(+1,+1)

	MV (Low)	MV (Medium)	MV (High)	BTM (Low)	BTM (Medium)	BTM (High)
Pos	0.0535**	0.0139	0.0018	-0.0074	0.0310	0.0512*
	(2.09)	(0.59)	(0.08)	(-0.34)	(1.30)	(1.90)
Neg	-0.0431**	-0.0127	-0.0107	-0.0089	-0.0187	-0.0366**
	(-2.53)	(-0.90)	(-0.94)	(-0.71)	(-1.56)	(-2.01)
Fund	-0.0005	0.0001	-0.0002	-0.0011	0.0016	-0.0008
	(-0.27)	(0.07)	(-0.15)	(-0.94)	(1.25)	(-0.50)
*Pos*Fund*	0.1834	0.1141	0.0386	0.1414*	0.0487	0.1207
	(1.60)	(1.29)	(0.55)	(1.84)	(0.62)	(0.99)
*Neg*Fund*	-0.0614	-0.0882*	-0.0215	-0.0098	-0.0813*	-0.0742
	(-1.05)	(-1.77)	(-0.60)	(-0.23)	(-1.79)	(-1.43)
MC	-0.0013	0.0004	-0.0021**	-0.0010	-0.0013	-0.0003
	(-1.10)	(0.38)	(-2.22)	(-0.91)	(-1.55)	(-0.30)
*Pos*MC*	0.0414	0.1256*	0.2853***	0.1389*	0.0750	0.1774**
	(0.51)	(1.75)	(4.10)	(1.82)	(1.16)	(2.29)
*Neg*MC*	0.0427	-0.0426	-0.0162	0.0123	0.0294	-0.0338
	(1.02)	(-1.17)	(-0.52)	(0.29)	(1.00)	(-0.93)
$FFCAR_{0,0}$	0.0016	0.0209	0.0096	0.0260**	-0.0122	0.0107
	(0.10)	(1.36)	(0.73)	(2.22)	(-0.79)	(0.59)
$FFCAR_{-1,-1}$	-0.0083	-0.0065	-0.0393***	-0.0278**	-0.0322***	-0.0029
	(-0.54)	(-0.41)	(-2.79)	(-2.40)	(-2.78)	(-0.15)
$FFCAR_{-2,-2}$	-0.0267*	-0.0129	-0.0366***	-0.0218*	-0.0488***	-0.0155
	(-1.80)	(-0.80)	(-2.88)	(-1.88)	(-3.94)	(-0.97)
$FFCAR_{-30,-3}$	-0.0016	-0.0021**	-0.0023***	-0.0015**	-0.0019**	-0.0026**
	(-1.53)	(-2.17)	(-2.76)	(-2.21)	(-2.37)	(-2.02)
$FFAlpha_{-252,-31}$	-0.0113***	-0.0082***	-0.0124***	-0.0105***	-0.0119***	-0.0117***
	(-5.05)	(-3.95)	(-6.11)	(-5.51)	(-5.90)	(-5.43)

Market Level Analysis

- Aggregate regression results

	FT	*ALL*
$AggPos_0$	0.0434***	0.0347**
	(3.0065)	(2.3104)
$AggNeg_0$	-0.0824***	-0.0575***
	(-5.5399)	(-3.9694)

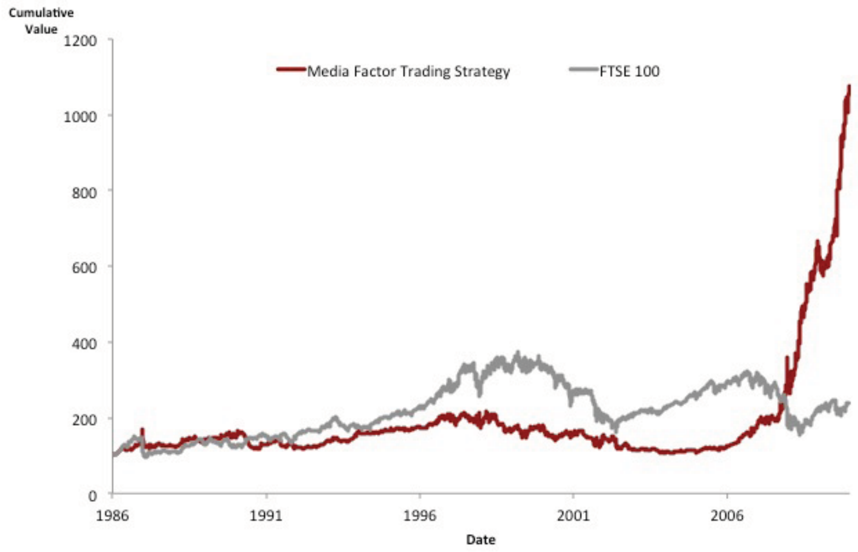
Trading Strategy I

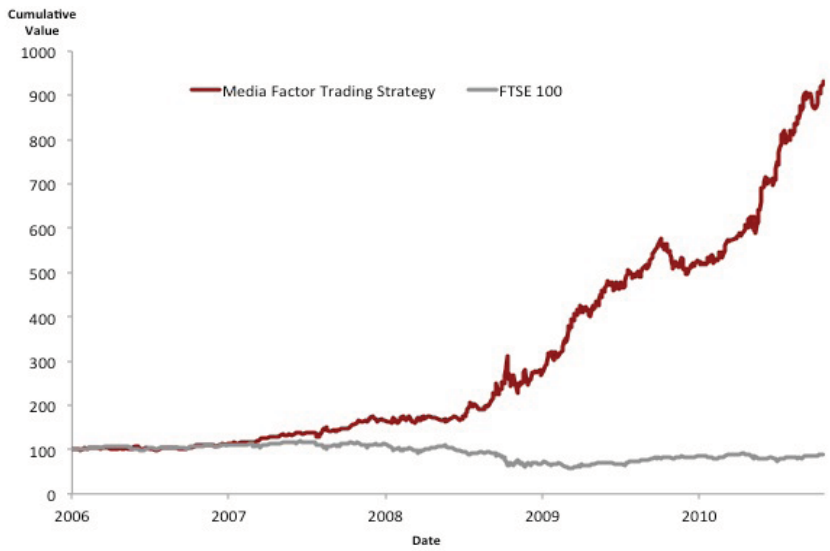
- Long short portfolio strategy. Equally weighted. With those stocks with Pos > Neg in long basket, with Pos < Neg in the short basket.

	FT				*ALL*			
	1987 -1994	1995 -2002	2003 -2010	1987 -2010	1987 -1994	1995 -2002	2003 -2010	1987 -2010
Alpha	-0.0210	0.1117	0.1903***	0.1164**	0.0304	0.0150	0.1419***	0.0676***
	(-0.23)	(1.25)	(2.61)	(2.38)	(0.95)	(0.38)	(3.85)	(3.21)
Market	-0.0544	-0.1177	-0.0449	-0.0591	-0.0041	0.0717	-0.0781	-0.0339
	(-0.35)	(-0.43)	(-0.42)	(-0.70)	(-0.07)	(0.65)	(-1.15)	(-0.83)
SMB	-0.0645	0.0198	-0.0153	-0.0056	-0.0216	0.0332	-0.0440	-0.0261
	(-0.63)	(0.14)	(-0.22)	(-0.11)	(-0.58)	(0.53)	(-0.87)	(-0.97)
HML	-0.0088	0.0652	0.0386	0.0533	0.0269	0.0293	-0.0386	-0.0011
	(-0.06)	(0.68)	(0.57)	(1.08)	(0.51)	(0.72)	(-0.90)	(-0.04)
UMD	-0.3332	-0.0763	-0.0727	-0.0819	0.1687*	0.1060*	-0.0013	0.0678*
	(-1.15)	(-0.49)	(-0.77)	(-1.05)	(1.91)	(1.69)	(-0.02)	(1.69)
Trading Days	467	946	1067	2485	1514	1843	1842	5229
Adjusted R^2	-0.0019	0.0021	-0.0009	0.0015	0.0025	0.0014	0.0010	0.0009

Trading Strategy II

- Trading strategy II uses an aggregate measures of positive and negative words across all FTSE 100 companies to derive a L/S signal on the FTSE 100 Index
- Positive words are inflated by 2.13
- Annual Sharpe of full sample strategy 0.28, annual excess return 5.15%, annual SD 18.30%
- Annual Sharpe of 2006-2010 strategy 2.36, annual excess return 55.24%, annual SD 23.41%. Significant alpha of 17bps per day (4-Factor).
- 2006-2010 correct prediction of FTSE direction 56%. For FTSE moves >1%, 62% correct.

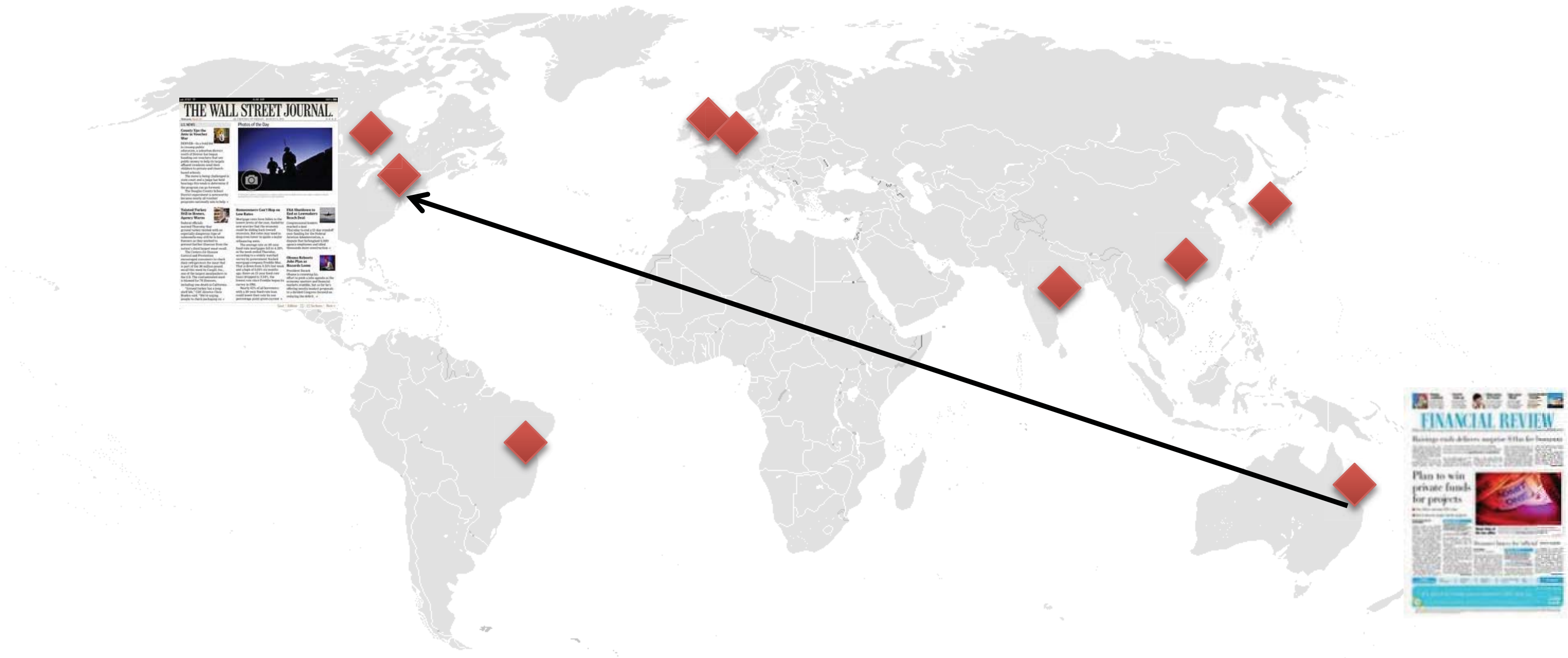




Paper II

Ferguson (2013) Investor Information Processing and Trading Volume. *National Stock Exchange of India – Student Research Paper*

Information Processing on a Global Scale



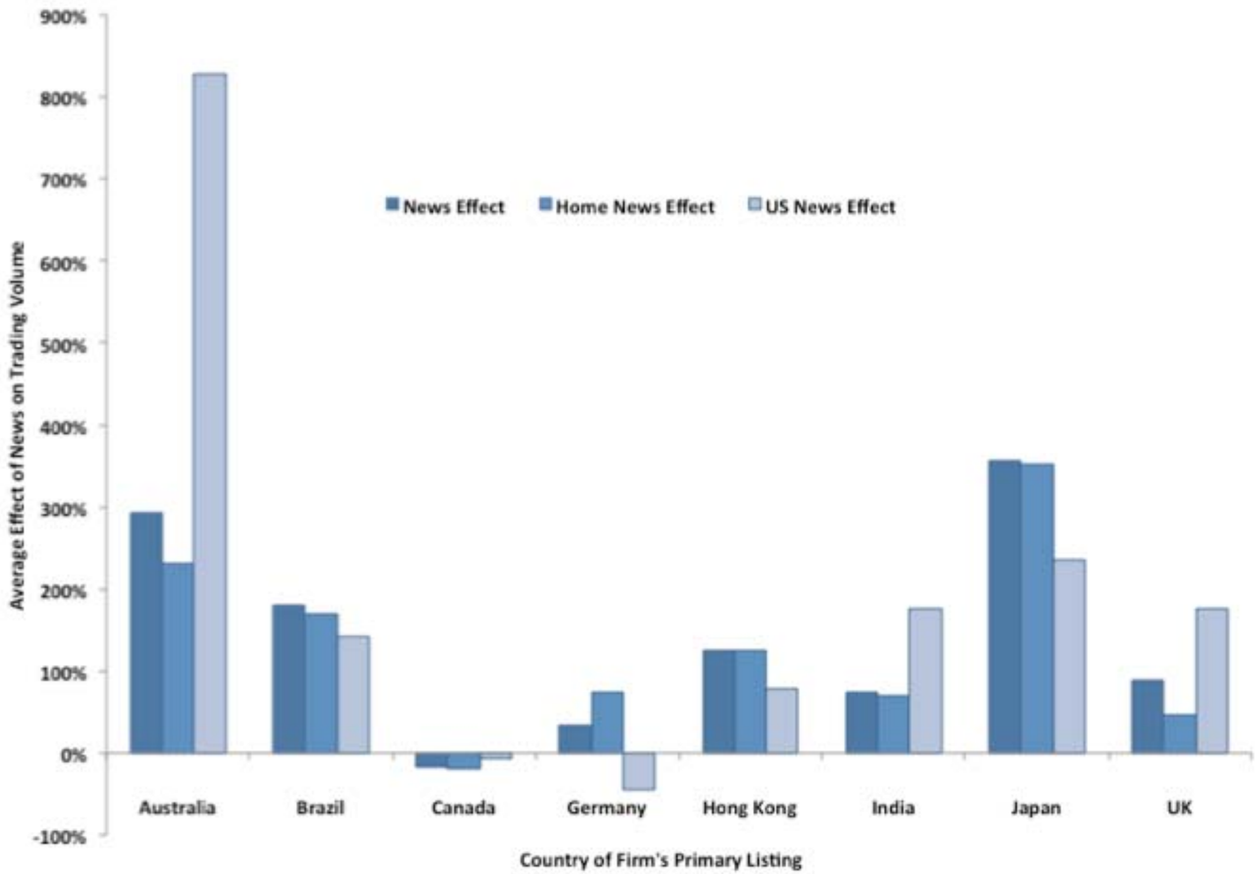
- Examine the effect of home market news and U.S. news on U.S. trading volume.
- NYSE largest exchange in the world, highest chances of secondary listing, strong governance practices, one of last to open on trading day.
- Sample introduces previously unconsidered heterogeneity in terms of language, distance, and visibility into firm specific news coverage.

Sample

Country	News Sources	Language
Australia	The Australian, Australian Financial Review	English
Brazil	Folha de S. Paulo, Valor Econômico	Portuguese
Canada	Globe and Mail, National Post (Financial Post)	English
Germany	Financial Times Deutschland, Handelsblatt	German
Hong Kong	Hong Kong Economic Times, Hong Kong Economic Journal	Chinese
India	Times of India, The Economic Times	English
Japan	Yomiuri Shimbun, Nihon Keizai Shimbun	Japanese
U.K.	Financial Times, The Times	English
U.S.	Wall Street Journal, New York Times	English

- 108 Firms
- 8 home market countries, on 5 continents, with 5 home market languages
- Sample covers 1997 to 2007
- 108,758 home news articles
- 19,948 U.S. news articles
- Average 9.93 articles per firm per month

News Effect on Trading Volume



- News – NO News, t-stat = **63.86**
- Home News – U.S. News, t-stat = **11.07**

Main Results

Regression analysis:- $\Delta Ln(Vol_{i,t}) = \beta_1 H_{i,t} + \beta_2 H_{i,t-1} + \beta_3 US_{i,t} + \beta_4 US_{i,t-1} + Controls$

- Trading volume rises in response to news on day *t*, and falls on day *t+1*
- Magnitude of reaction to U.S. news significantly greater than to home news (F, 7.60)
- Salience of U.S. news much greater to U.S. based investors

Independent Variables	Full Sample (1)	
$H_{i,t}$	0.0062	***
	(0.00)	
$H_{i,t-1}$	-0.0038	***
	(0.00)	
$US_{i.t}$	0.0164	***
	(0.01)	
$US_{i,t-1}$	-0.0066	***

- Proxies for positive and negative news indicate relationship between trading volume and news driven by positive news.

D=1 if $H_{i,t}\|US_{i,t}>0$ & $R_{i,t}>0$	0.0045	
	(0.01)	
D=1 if $H_{i,t}\|US_{i,t}>0$ & $R_{i,t}<0$	-0.0202	***
	(0.01)	

Results - Language

Independent Variables	English		Non-English	
$H_{i,t}$	0.0053	***	0.0059	***
	(0.00)		(0.00)	
$H_{i,t-1}$	-0.0043	***	-0.0029	***
	(0.00)		(0.00)	
$US_{i.t}$	0.0248	***	0.0062	
	(0.01)		(0.00)	
$US_{i,t-1}$	-0.0055	*	-0.0065	

- English language countries:- Australia, Canada, India, & U.K.
- Non-English language countries:- Brazil, Germany, Hong Kong, & Japan

- U.S. news has no significant impact on trading volume of firms from non-English language countries
- Translation and search costs, information asymmetries, U.S. news sources may not be able to retrieve as much value relevant information, hence lower demand

Results - Visibility

Independent Variables	High		Medium		Low	
$H_{i,t}$	0.0068	***	0.0058	*	0.0047	**
	(0.00)		(0.00)		(0.00)	
$H_{i,t-1}$	-0.0018	**	-0.0044		-0.0061	***
	(0.00)		(0.00)		(0.00)	
$US_{i,t}$	0.0110	**	0.0362	***	0.0125	
	(0.01)		(0.01)		(0.03)	
$US_{i,t-1}$	-0.0099	***	0.0067		0.0002	
	(0.00)		(0.01)		(0.01)	

- Sample split equally by average analyst coverage to proxy for information asymmetries

- Firms with low levels of analyst coverage are not sensitive to U.S. news.
- For firms with greater information asymmetries, information retrieval is more costly due to lower demand for information.

Summary and Conclusions

- Textual analysis can be made easy with the use of dictionaries and tagged data sources. Relatively straightforward to code and implement.
- News media play a important role in delivering value relevant information to investors.
- Measures of semantics in news articles can predict next day stock returns, and aggregate measures can predict next day market returns.
- The effect of semantics on stock returns is higher for smaller firms with greater informational risk.
- Investor dependence on home market information increases for firms with greater information asymmetries. Analyst coverage is an important aspect of firms' information flow such that it influences investor demand for alternative sources of information.
- Characteristics of information may present interesting investment opportunities.