When can the market identify old news?

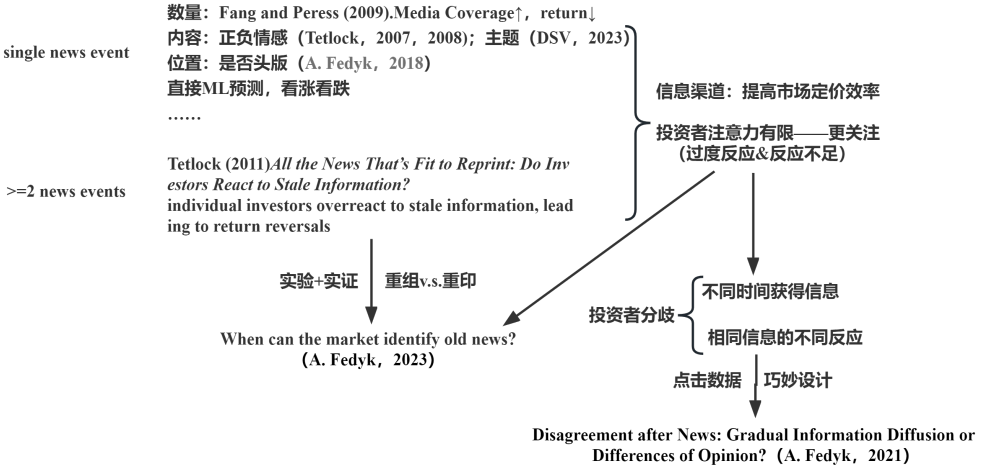
Anastassia Fedyk, James Hodson(JFE, 2023)

石宛青

（武汉大学金融系）

2022 年 11 月 7 日

Framework


single news event
数量：Fang and Peress (2009).Media Coverage↑，return↓
内容：正负情感（Tetlock，2007，2008）；主题（DSV，2023）
位置：是否头版（A. Fedyk，2018）
直接ML预测，看涨看跌
……
>=2 news events
Tetlock (2011)All the News That's Fit to Reprint: Do Investors React to Stale Information?
individual investors overreact to stale information, leading to return reversals
信息渠道：提高市场定价效率
投资者注意力有限——更关注
（过度反应&反应不足）
实验+实证
重组v.s.重印
When can the market identify old news?
（A. Fedyk，2023）
投资者分歧
不同时间获得信息
相同信息的不同反应
点击数据
巧妙设计
Disagreement after News: Gradual Information Diffusion or Differences of Opinion?（A. Fedyk，2021）

Motivation

- Is "news" new? Not yet
- voluminous news V.S.limited attention – difficult to identify new content
- puzzle of market reactions to old news(Tetlock,2011； Gilbert et al. ,2012)
  - eg:A front page article in the New York Times in 1998.5, which largely repeated information from five months prior, prompted 330% price increase for the firm.
- an idea: specific limited attention-correlation neglect-"recombination effect"

Question

- Can investor identify old news(reprint&recombination) when it draws from multiple previous articles?
  - an experiment on 155 active finance professionals
- Pricing "recombination effect"?
  - test the asset pricing using a unique dataset of all financial news appearing on the Bloomberg terminal.

Conclusion

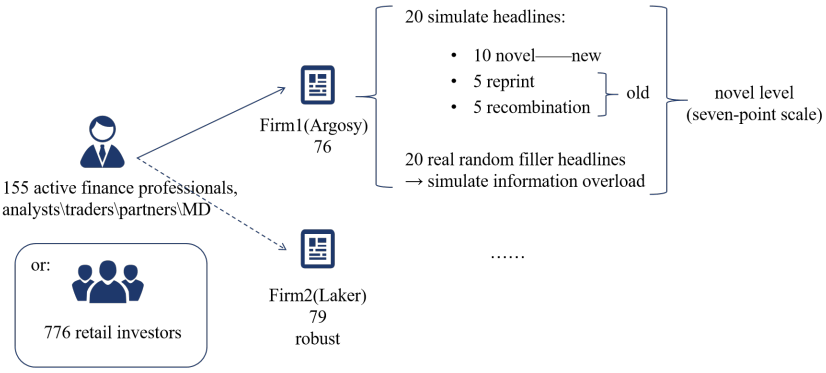
- even sophisticated investors: articles that combine information from several places are more difficult to distinguish from novel information.

- market reactions are significantly larger in response to recombination articles than in response to simple reprints, and reverse during the following week.

Contribution

- Extend Tetlock (2011) from the perspective of divided old news into reprint and recombination, get more accurate results: investors cannot identify recombination
- Innovative method of calculating text similarity, different from Tetlock (2011)
- Regarding the reaction of inst investors: Tetlock (2011) believes that inst do not overreact. This paper use experiment and empirical prove that even inst overreact to recombination.

Experiment



- novel: 20% words appeared in previous headlines about the same firm.
- old news: 80%. Reprints: comes from a single preceding headline about the same firm; Recombinations: at least two prior headlines.

Experiment

- Example:
  1. Argosy's misfit design business down, some tough questions to answer
  2. Argosy Trucking Q3 results above expectations eps 1.2 vs 1.1
  3. Argosy beats expectations: Q3 trucking results eps up 0.1 on 1.1
  4. Argosy Q3 earnings beat expectations, but design business down

- 3 is a reprint of 2, with 75% of its words already appearing previously

- 4 is a recombination: 82.5% of its words appearing previously, but it combines equal parts from 1 and 2.

Experiment

- Example:
  1. Argosy's misfit design business down, some tough questions to answer
  2. Argosy Trucking Q3 results above expectations eps 1.2 vs 1.1
  3. Argosy beats expectations: Q3 trucking results eps up 0.1 on 1.1
  4. Argosy Q3 earnings beat expectations, but design business down

- 3 is a reprint of 2, with 75% of its words already appearing previously

- 4 is a recombination: 82.5% of its words appearing previously, but it combines equal parts from 1 and 2.

Experiment

Panel 1: Length (# words)

	Novel	Reprint	Recombination
Argosy sample	7.8	7.8	7.8
Laker sample	7.6	7.4	7.4
Overall	7.7	7.6	7.6

Panel 2: Location (#1-#20)

	Novel	Reprint	Recombination
Argosy sample	9.1	12.6	11.2
Laker sample	8.8	11.6	12.8
Overall	9.0	12.1	12.0

Panel 3: Old content (% words seen)

	Novel	Reprint	Recombination
Argosy sample	22.2%	82.7%	81.1%
Laker sample	20.8%	79.4%	81.1%
Overall	21.5%	81.1%	81.1%

Test Asset pricing-Prediction

1. Compared to novel news, old news is associated with lower trading volumes and absolute price changes immediately following publication.
2. Among old news, recombination articles are associated with larger trading volumes and absolute price changes than reprint articles.
3. Initial reactions to old news are subject to subsequent reversal. In particular, during the days or weeks following news publication:
   - The initial price moves after old news see more reversal than the initial price moves after novel news.
   - The initial price moves after recombination articles see more reversal than the initial price moves after reprint articles.

Test Asset pricing-Data

- Widely used Factiva, Dow Jones' database: smaller and less real-time
- Bloomberg, one of the most comprehensive,within 100 milliseconds
  - news written and published by Bloomberg directly (10%);
  - national and international news from partner news organizations (60%);
  - web sources, including regional and local news, blogs, and social media (30%).
- conditions for limiting noise:
  - tagged with security codes traded in U.S.
  - exclude stock <$5 per share ——29,500 news per day.
  - at least 70% relevance score for at least one U.S.-traded equity security.
    - 90% are highly targeted to the tagged security(earnings, products, or strategy);
    - 70% are less but still relevant(eg:main competitor);
    - 50% are only tangentially relevant for the security in question.
  - final: 4000 news per day, each news linked to an average of 1.3 securities.

Test Asset pricing-Data

- eg:AAPL-relevance score
  - “Apple announces event on 3/17 to unveil new iPad” (95% relevance)
  - “Android Grows U.S. Smartphone Market Shr to 50.1%” (70% relevance)
  - “JCPenney lowered to BB at Standard & Poor’ s on new strategy” (50% relevance)

Test Asset pricing-Measure

- Old news, for each news s:
  - extract the unique words (unigrams)(exclude stop words; "earned", "earnings" to "earn-")
  - number of unique term:$||\cdot||$; $||s_1 \cdot s_2||$ appearing in both s1 and s2

$$Ols(s) = \frac{||s \cap (\cup_{i=1}^{5} s_i'(s))||}{||s||} \tag{1}$$

- s_i' stock i published up to three days (72 hours) before the publication of s
- $(s_1', ..., s_5')$ individually span the largest fraction of terms in s

Test Asset pricing-Measure

- Old news, for each news s:

$$Ols(s) = \frac{||s \cap (\cup_{i=1}^{5} s_i'(s))||}{||s||} \tag{1}$$

- key innovation: V.S. Tetlock (2011) average intersection

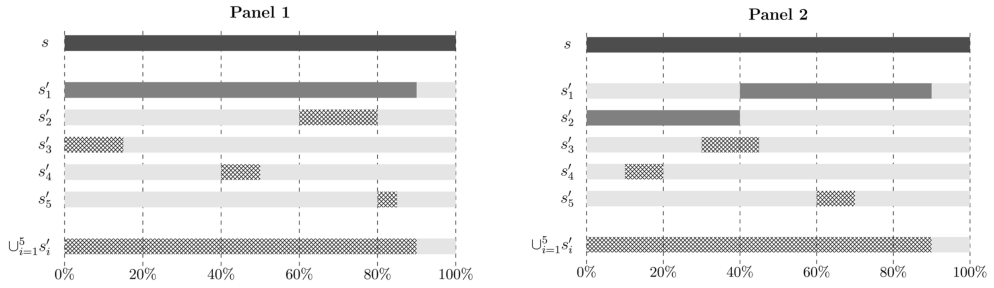
s	s'	s''	Totlock(2011)	this paper
AB	AC	AD	50%	50%
AB	AB	CD	50%	100%

- Robust: using bigrams (pairs of words)
- old news definition: at least 60% of $Ols(s)$

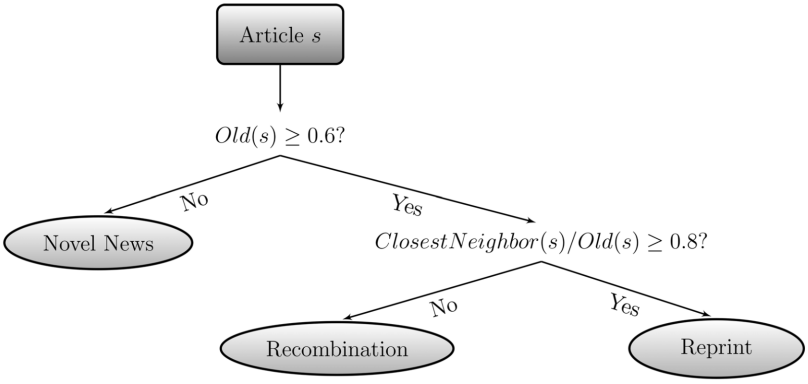
Test Asset pricing-Measure

- Old news: Reprints and recombinations, for each news s:

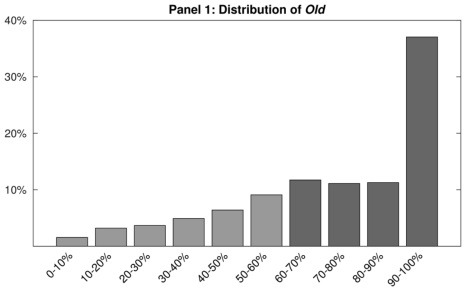
$$ClosestNeighbor(s) = \frac{max_{s'}||s \cap s'||}{||s||} = \frac{||s \cap s'_1(s)||}{||s||} \quad (2)$$

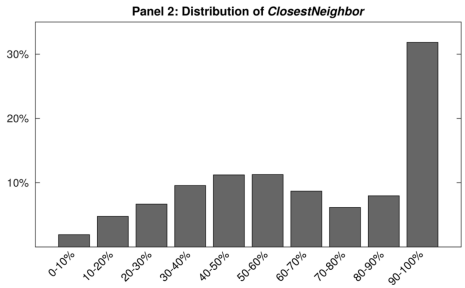


Test Asset pricing-Measure



Test Asset pricing-Measure





- Nnearly 40%, are almost entirely (90% or more) spanned by preceding news of same firm.
- 40% articles have appeared(80% or more) in the single closest neighbor (capturing exact reprints),others are recombinations.

Test Asset pricing-Measure

- aggregate individual article-level classifications to the firm-day level.

	Article-Level			Firm-Level		
	# Terms	Old	ClosestNeighbor	# Terms	% Old News	% Recombinations
Full Sample						
Mean	146	0.72	0.52	151	72%	21%
Median	138	0.77	0.56	146	69%	20%
25% Percentile	67	0.65	0.43	110	56%	14%
75% Percentile	201	0.94	0.92	173	82%	25%
By year (mean)						
2000	200	0.68	0.43	200	70%	23%
2001	200	0.68	0.43	199	71%	24%
2002	202	0.68	0.44	200	70%	21%
2003	194	0.68	0.43	188	69%	22%
2004	196	0.70	0.48	190	72%	23%
2005	202	0.69	0.46	196	72%	19%
2006	157	0.71	0.50	158	74%	20%
2007	161	0.72	0.52	160	75%	18%
2008	155	0.68	0.46	151	69%	18%
2009	133	0.67	0.47	129	68%	15%
2010	130	0.67	0.45	135	68%	16%
2011	120	0.72	0.54	130	71%	18%
2012	129	0.70	0.51	129	70%	17%
2013	138	0.79	0.54	144	75%	19%
2014	138	0.80	0.55	142	72%	20%

Test Asset pricing-Model

- prediction1: old news-lower trading vol and abs price changes immediately
- prediction2: recombination-Larger trading vol and abs price changes immediately

$$|AbnRet|_{i,t} = a + b_1 AbnPrcOld_{i,t} + b_2 AbnPrcRecombination_{i,t} + gX_i, t + e_{i,t} \quad (3)$$

$$AbnVol_{i,t} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \quad (4)$$

- $|AbnRet|_{i,t}$: abs of dif between i's ret and value-weighted ret on t
- $AbnVol_{i,t}$: dif between i's fraction of shares turned over and VW on t

Test Asset pricing-Model

- prediction3: old news、recombination see more reversal

$$\begin{aligned} AbnRet_{i,[t+t_1,t+t_2]} = \alpha &+ \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcOld_{i,t} \times AbnRet_{i,t} \\ &+ \beta_3 AbnRet_{i,t} + \sigma_1 AbnPrcRecombination_{i,t} \\ &+ \sigma_2 AbnPrcRecombination_{i,t} \times AbnRet_{i,t} \\ &+ \gamma X_{i,t} + \epsilon_{i,t} \end{aligned} \tag{5}$$

- σ_2 measures differential reversal following larger shares of recombination news –negative

Experiment result

Panel 1: Mean responses

	Novel	Reprint	Recombination
7-point scale	4.52	2.61	3.03
5-point scale	3.84	2.40	2.63

Panel 2: Comparisons

		Novel vs. Old	Recombination vs. Reprint
7-point scale	Diff	1.70***	0.42**
	(SE)	*(0.15)*	*(0.20)*
5-point scale	Diff	1.33***	0.23**
	(SE)	*(0.06)*	*(0.10)*

- participants can identify novel news as more new information than old news.
- participants are more susceptible to recombination than reprinting.

Empirical result

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		Dependent variable: $AbnVol_{i,t}$	
	(1) Old News Only	(2) Old News & Recombinations	(3) Old News Only	(4) Old News & Recombinations
$AbnPrcOld_{i,t}$	-0.098%***	-0.115%***	-0.062%***	-0.075%***
	(0.017%)	*(0.019%)*	*(0.006%)*	*(0.007%)*
$AbnPrcRecombination_{i,t}$		0.176%***		0.088%***
		(0.016%)		*(0.006%)*
Controls:				
$Stories_{i,t}$	X	X	X	X
$AbnStories_{i,[t-5,t-1]}$	X	X	X	X
$Terms_{i,t}$	X	X	X	X
$MCap_{i,t}$	X	X	X	X
$BM_{i,t}$	X	X	X	X
$AbnRet_{i,[t-5,t-1]}$	X	X	X	X
$AbnVol_{i,[t-5,t-1]}$	X	X	X	X
$AbnVolatility_{i,[t-5,t-1]}$	X	X	X	X
$Illiq_{i,[t-5,t-1]}$	X	X	X	X
R^2	0.243	0.252	0.185	0.197

- an additional 10% of the news about firm *i* being old corresponds to an 11 bps-smaller abs abn ret and 0.08% lower abn vol.

Empirical result

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		Dependent variable: $AbnVol_{i,t}$	
	(1) Old News Only	(2) Old News & Recombinations	(3) Old News Only	(4) Old News & Recombinations
$AbnPrcOld_{i,t}$	-0.098%***	-0.115%***	-0.062%***	-0.075%***
	(0.017%)	*(0.019%)*	*(0.006%)*	*(0.007%)*
$AbnPrcRecombination_{i,t}$		0.176%***		0.088%***
		(0.016%)		*(0.006%)*

- an additional 10% of recombination old news corresponds to an additional 18bps abs abn ret and an additional 0.09% abn vol.
- if a firm has an additional 10% of recombination news on a given day, it experiences on average a 6 bps larger abs abn ret (0.176%−0.115% = 0.061%).

Empirical result

- recombination's market reacts more strong. But do these stronger responses reflect overreactions?
- if this reactions are warranted, then we should observe no subsequent reversal.

	(1) $AbnRet_{i,[t+2,t+5]}$	(2) $AbnRet_{i,[t+1,t+5]}$	(3) $AbnRet_{i,[t+2,t+10]}$
$AbnRet_{i,t}$	-0.044**	-0.051**	-0.039
	(0.020)	*(0.023)*	*(0.033)*
$AbnPrcOld_{i,t} * AbnRet_{i,t}$	-0.094***	-0.139***	-0.168*
	(0.025)	*(0.029)*	*(0.038)*
$AbnPrcRecombination_{i,t} * AbnRet_{i,t}$	-0.131***	-0.169***	-0.199**
	(0.039)	*(0.059)*	*(0.098)*
R^2	0.092	0.098	0.082

- an additional 10% old news of firm$i$, then next day ret: a thrice larger reversal
- the additional reversal in response to recombination news is even starker,

The role of investor attention

- Data: "News Heat –DailyMax Readership" :based on the daily maximum of eight-hour news reads and searches for a given security.

Panel 1: Market Reaction

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		Dependent variable: $AbnVol_{i,t}$	
	(1) Low attention	(2) High attention	(3) Low attention	(4) High attention
$AbnPrcOld_{i,t}$	-0.086%***	-0.159%***	-0.004%	-0.123%***
	(0.020%)	*(0.027%)*	*(0.007%)*	*(0.008%)*
$AbnPrcRecombination_{i,t}$	0.164%***	0.180%***	0.028***	0.102%***
	(0.019%)	*(0.025%)*	*(0.007%)*	*(0.008%)*

Panel 2: Return Reversal

	Dependent variable: $AbnRet_{i,[t+2,t+5]}$	
	Low attention	High attention
$AbnExtentOld_{i,t} \times AbnRet_{i,t}$	-0.105***	-0.056
	(0.038)	*(0.049)*
$AbnExtentRecombination_{i,t} \times AbnRet_{i,t}$	-0.133**	-0.099*
	(0.052)	*(0.059)*

- high attention-help to lower old news reaction; but not by recombinations

Retail and institutional investors

- retail investors:more cognitive biases, including overreactions to old news? (investor ownership from NASDAQ)

Panel 1: Market Reaction

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		*Dependent variable:* $AbnVol_{i,t}$	
	(1) Low instit. ownership	(2) High instit. ownership	(3) Low instit. ownership	(4) High instit. ownership
$AbnPrcOld_{i,t}$	-0.076%***	-0.123%***	-0.002%	-0.077%***
	(0.022%)	*(0.020%)*	*(0.005%)*	*(0.013%)*
$AbnPrcRecombination_{i,t}$	0.168%***	0.175%***	0.079%***	0.055%***
	(0.024%)	*(0.023%)*	*(0.008%)*	*(0.010%)*

Panel 2: Return Reversal

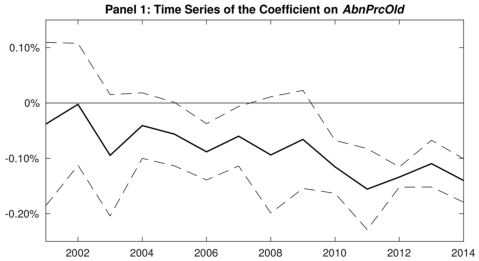
	Dependent variable: $AbnRet_{i,[t+2,t+5]}$	
	Low institutional ownership	High institutional ownership
$AbnExtentOld_{i,t} \times AbnRet_{i,t}$	-0.104**	-0.085**
	(0.042)	*(0.040)*
$AbnExtentRecombination_{i,t} \times AbnRet_{i,t}$	-0.112**	-0.145***
	(0.049)	*(0.051)*

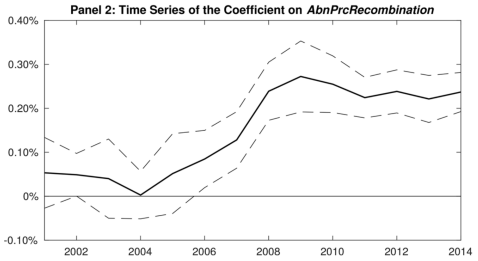
- intitutional investors are better at screening out old news but still can't identify recombinations

News sentiment and ambiguity

- methodology for measuring news sentiment and ambiguity follows Fedyk (2021).
  - Human experts manually tag 10,000 news articles as positive, negative, or neutral, and as hard or soft information.
  - represented as vectors of features including length, topics, indicators for specific unigrams···
  - use SVM to classify other news articles based on the attributes learned from the training data.
  - The ambiguity score is an average of two components:
    - the confidence with which the method identifies the article' s sentiment
    - whether the article is classified as hard information, interacted with the confidence of this classification.
- No effect

Time series of the effects





- Market reactions to old news, compared to new news, declined from 2001-2014.
- By contrast, the differential reaction to recombination increased over time.
  - algorithmic tools aim to identify with novel news. However, they prioritize speed.

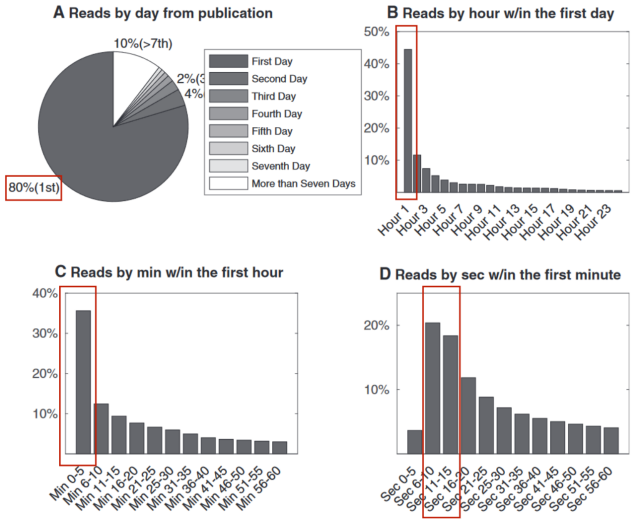
Robustness

- Instead of using simple abnormal returns, we compute characteristics-adjusted returns following the methodology in Daniel et al. (1997).
- measures old news and recombinations consider:
  - the look-back window (π) to five and 10 previous business days;
  - the number (n) of considered most similar articles about the same firm to 10
  - consider firm-dates that actually have at least n articles in the preceding π days.
- alternative approach based on each article's continuous measures of old and recombined content
- reclassify articles into novel news, reprints, and recombinations using an old content threshold of 90%, rather than 60%.
- Robust

Disagreement after News: Gradual Information Diffusion or Differences of Opinion?

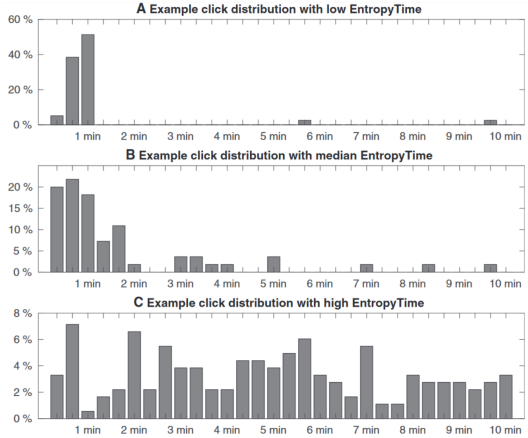
- Fedyk,2023,Review of Asset Pricing Studies
- news can spur increased trading volume-channel:disagreement(Hong and Stein,2007)
- where disagreement from?
  - investors acquiring information at different times
  - investors reacting differently to the same information
- Data: clicks by finance professionals on 3.5 million news articles between March 2014 and March 2015.
  - individual clicks, can observe dynamics of investor attention at high frequency
  - although data are anonymized, clicks by the same reader are linked to each other, allow to classify readers into types

Disagreement after News——Design



Disagreement after News——Design

- Measuring gradual information diffusion
- normalized Shannon entropy of read times (Shannon 1948).



Disagreement after News——Design

- Measuring differences of opinion
  - encoding readers according to their clicks on every news article would yield a very large (and sparse) matrix.
  - **Industry focus (23 features)**. For each of the 23 two-digit NAICS industry codes, I consider whether news about that industry captures a meaningful share (at least 5%) of a given reader' s clicks.
  - **Firm-level focus (3 features)** I consider three aspects of firm-specific focus: whether a given reader follows an unusually high (or unusually low) number of securities and whether the reader has an especially strong preference for a single firm.
  - News source focus (3 features);News source types (16 features);Reading gaps (3 features);Length of stories read (2 features);Reading of stale and reprint stories (10 features);Activity level (6 features)

Disagreement after News——Design

- Measuring differences of opinion
  - After 66-dimensional feature space, cluster them into types using affinity propagation, an unsupervised learning technique proposed by Frey Dueck (2007)
  - identifies 21 fairly balanced clusters with 100–300 points in each
  - **Specialist reader**: disproportionately follows a single news source, prefers short stories, follows a single industry, and has historically been moderately active;
  - **Generalist reader**: broad source focus, has very few long lags between reads, prefers short stories, and has a broad firm focus;

Disagreement after News——Result

- over 48 hours around earnings announcements (panel A) and over 10 minutes following individual news articles (panel B)——low correlations

Table 1
Descriptive statistics of the key empirical variables

A. Earnings announcement setting

	Mean	Median	SD	Min	10%	25%	75%	90%	Max
EntropyTime	0.51	0.51	0.10	0.05	0.38	0.45	0.57	0.63	0.94
EntropyType	0.69	0.70	0.08	0.14	0.50	0.65	0.75	0.78	0.88
AbnVolume	1.77	0.86	10.00	−0.54	−0.04	0.31	1.75	3.22	12.38

Correlations with:

	EntropyTime	EntropyType	Total Reads	Size	B/M	Absolute SUE
EntropyTime	–	0.15	−0.03	−0.01	−0.02	0.03
EntropyType	0.15	–	0.34	0.31	0.29	0.00

B. High-frequency analysis around individual news events

	Mean	Median	SD	Min	10%	25%	75%	90%	Max
EntropyTime	0.65	0.64	0.14	0.32	0.49	0.56	0.75	0.85	0.96
EntropyType	0.56	0.57	0.12	0.34	0.40	0.48	0.64	0.70	0.81
AbnVolume	0.18	0.07	1.02	−0.93	−0.57	−0.35	0.53	1.02	6.03

Correlations with:

	EntTime	EntType	Total Reads	Size	B/M	Stale	Sent
EntropyTime	–	0.49	0.46	−0.06	0.01	−0.20	−0.15
EntropyType	0.49	–	0.24	0.09	0.03	0.01	−0.02

Disagreement after News——Result

Table 3

Trading volume tests around earnings announcements

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EntropyTime	1.60***	1.55***	1.53***	1.69***	1.64***	1.61***			
	(0.38)	(0.39)	(0.40)	(0.37)	(0.39)	(0.39)			
EntropyType	0.55*	0.52	0.67**				0.66**	0.69**	0.84***
	(0.32)	(0.33)	(0.34)				(0.31)	(0.31)	(0.32)
TotalReads	X	X	X	X	X	X	X	X	X
Size	X	X	X	X	X	X	X	X	X
B/M	X	X	X	X	X	X	X	X	X
Absolute SUE	X	X	X	X	X	X	X	X	X
Year FE		X	X		X	X		X	X
Day-of-week FE		X	X		X	X		X	X
Industry FE			X			X			X
*R*2	.11	.11	.14	.08	.09	.12	.07	.07	.10
Obs.	8,337	8,337	8,337	8,337	8,337	8,337	8,337	8,337	8,337

- both channels are at play and that neither, on its own, offers a complete picture of disagreement and trading volume around information releases.

Disagreement after News——Result

Table 4
Trading volume tests around individual news articles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EntropyTime	4.10***	4.32***	4.23**	4.74***	4.79***	4.83***			
	(1.53*)	(1.32)	(2.13)	(1.09)	(1.14)	(1.64)			
EntropyType	2.85**	2.26**	2.58**				2.89***	2.34**	2.48***
	(1.03)	(1.07)	(1.10)				(0.88)	(0.95)	(0.96)
TotalReads	X	X	X	X	X	X	X	X	X
Size	X	X	X	X	X	X	X	X	X
B/M	X	X	X	X	X	X	X	X	X
Staleness	X	X	X	X	X	X	X	X	X
Sentiment	X	X	X	X	X	X	X	X	X
Source FE		X	X		X	X		X	X
Date FE		X	X		X	X		X	X
Hour FE		X	X		X	X		X	X
Firm FE			X			X			X
$R2$	.17	.19	.24	.15	.17	.23	.15	.18	.22
Obs.	226,641	226,641	226,641	226,641	226,641	226,641	226,641	226,641	226,641

- going from completely concentrated to maximally dispersed timing of clicks corresponds to an additional 423% increase in trading volume
- going from fully concentrated to fully dispersed types of readers attending to a given news corresponds to an additional 258% increase in short-term vol.

Highlight

- click data
- good design two channels for investors to obtain information at different times and react differently to the same information

Future extension

- 还可以从新闻、文本中挖掘哪些投资者注意力有限的故事、渠道?
- 考虑各类投资者注意力问题，优化新闻文本因子
- click 数据还有哪些应用？（EDGAR）

Thanks!