Discussion of:
Fake News: Evidence from Financial Markets

Shimon Kogan, Tobias Moskowitz and Marina Niessner

Kent Daniel[†]

[†]Columbia Business School & NBER

NBER Asset Pricing Summer Institute
July 12, 2019

Columbia Business School

Seeking Alpha

- Seeking Alpha is an online platform founded in 2004.
  - As of 2014, there were about 250 new articles/day and about nine million unique visitors a month.
  - Seeking Alpha authors are thousands "...of self-directed investors and other students of the market, many of whom choose to use pseudonyms."
  - There is almost no content control.[1]
- Chen, De, Hu, and Hwang (2014), "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media" examine SA data from 2005-2012, and find that:

  > *...the views expressed in both articles and commentaries predict future stock returns and earnings surprises.*

[1] See https://blogs.wsj.com/venturecapital/2014/03/19/study-crowdsourced-stock-opinions-beat-analysts-news/.

Seeking Alpha

- Seeking Alpha is an online platform founded in 2004.
  - As of 2014, there were about 250 new articles/day and about nine million unique visitors a month.
  - Seeking Alpha authors are thousands "...of self-directed investors and other students of the market, many of whom choose to use pseudonyms."
  - There is almost no content control.[1]
- Chen, De, Hu, and Hwang (2014), "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media" examine SA data from 2005-2012, and find that:

  > *...the views expressed in both articles and commentaries predict future stock returns and earnings surprises.*

[1] See https://blogs.wsj.com/venturecapital/2014/03/19/study-crowdsourced-stock-opinions-beat-analysts-news/.

Rick Pearson

- Rick Pearson is a former investment banker, a private investor, and a regular contributor to *Seeking Alpha*.
- In 2014 Pearson was approached by an investment-relations firm, "Dream Team", employed by firms looking for "good press". Pearson states:

  > *I was asked to write paid promotional articles on Galena Biopharma and CytRx Corp. (CYTR), without disclosing payment.*

- Pearson instead decided to short shares of CytRX, went undercover so as to dug deeper into this practice, and informed the SEC.[2]
  - Pearson found numerous other SA authors who had been paid for stories.

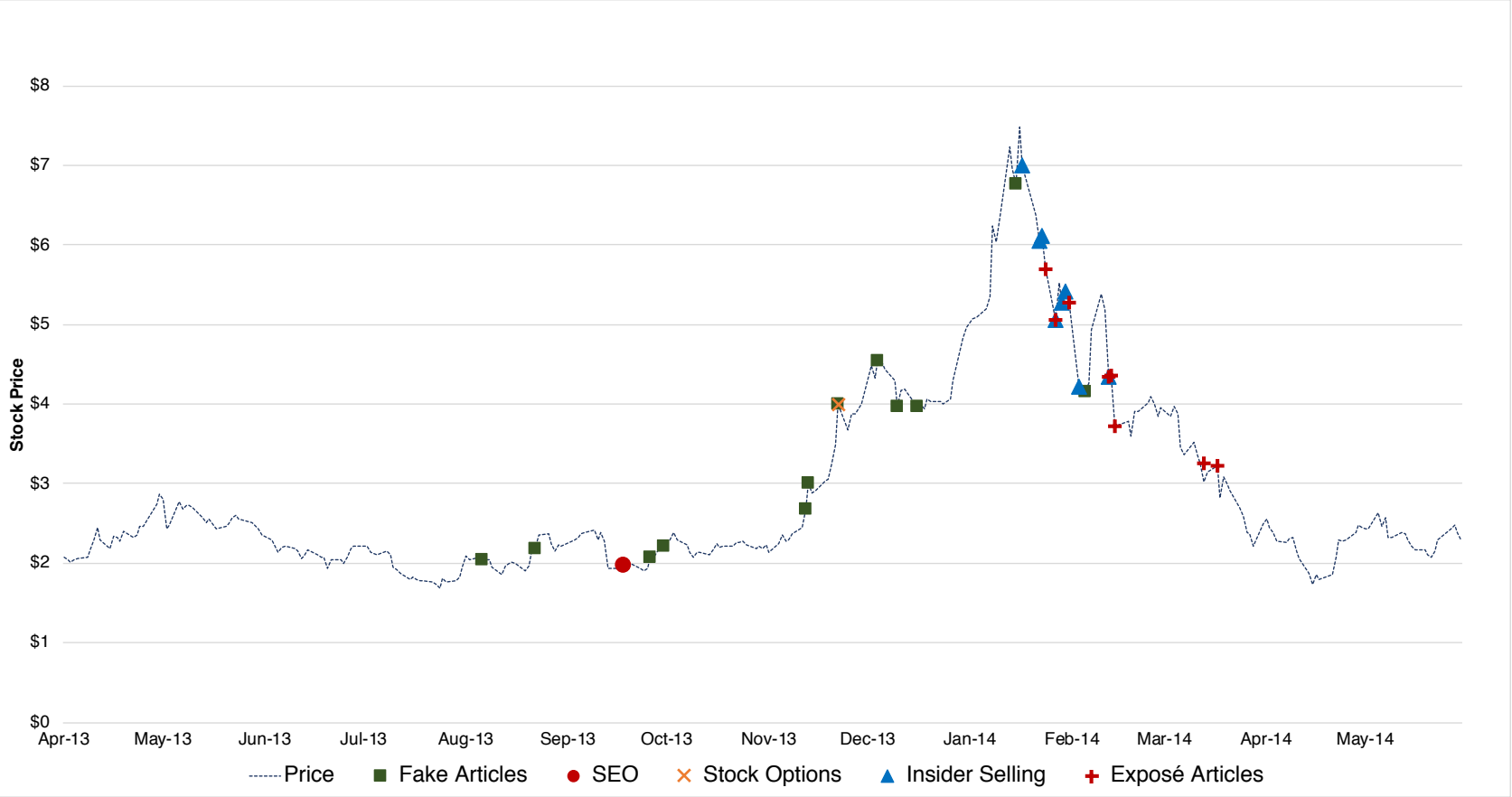
[2] See https://seekingalpha.com/article/2086173-behind-the-scenes-with-dream-team-cytrx-and-galena and https://www.barrons.com/articles/seeking-alpha-needs-to-take-stock-of-its-policies-1395420277.

Rick Pearson

- After being approached by the SEC, *Seeking Alpha* removed the articles in March 2014.
- Starting in October 2014, the SEC initiated lawsuits against the some of authors, the promotional firms, and the companies and executives who hired the promotional firms
  - At least some of these articles were related to "pump-and-dump" schemes.
- Pearson provided KMN with the 111 articles he found to be fake, and *Seeking Alpha* shared another 147 "fake" articles.
  - The authors match these articles to 60 exchange-traded firms.
- *Note:* In March 2017, Cemtrex filed a lawsuit against Pearson and three others alleging trade libel and tortious interference, but this suit was dismissed "without prejudice"in June 2017.[3]
  - From December 2016 to December 2018, Cemtrex fell from $7/shr to below $1/share.

[3] See https://www.businesswire.com/news/home/20170306006247/en/Cemtrex-Files-170-Million-Lawsuit-Richard-Pearson and https://seekingalpha.com/instablog/15663412-joe_retail/5246218-vuzix-following-cemtrexs-lead

Galena Biopharma, Inc.



For-sure fake- and non-fake-articles

- KMN end up with a sample of 171 "for-sure" fake articles, by 20 authors, about 47 exchange-traded firms.
- The authors argue that these articles are "intended to deceive"
- They also collect 334 additional articles from the same authors that are "not fake."
- This provides a *training sample* for the classification algorithm.
- The classification algorithm is then applied to a *test sample*:
  - The test sample is a set of ∼350,000 articles from *Seeking Alpha* and *Motley Fool*, classified as "fake" and "not-fake" articles.

Key Hypothesis

Fake News (FN) is different from real news.

- FN is intended to deceive.
  - FN authors are paid agents acting on behalf of some principal.
  - The principal's goal is price manipulation.
- FN authors (unconsciously?) change the way they write when they are trying to deceive
  - The word choices/grammar/syntax/sentence length change.
  - NLP methods can pick these changes up.
  - Presumably the authors (and the principals) are not aware of these changes.

Key Hypothesis

Fake News (FN) is different from real news.

- FN is intended to deceive.
  - FN authors are paid agents acting on behalf of some principal.
  - The principal's goal is price manipulation.
- FN authors (unconsciously?) change the way they write when they are trying to deceive
  - The word choices/grammar/syntax/sentence length change.
  - NLP methods can pick these changes up.
  - Presumably the authors (and the principals) are not aware of these changes.

The LIWC Authenticity Score

- KMN use the publicly available LIWC *Linquistic Inquiry and Word Count* software.
  - Pennebaker, Booth, and Francis (2007)
  - Authenticity classification (0-100) is based on word choices and counts.
  - Cecchini, Aytug, Koehler, and Pathak (2010), "Making Words Work: Using financial text as a predictor of financial events", use the LIWC to (successfully) forecast fraud and bankruptcy.
- For this training sample:
  - the fake articles have a average authenticity score of 19.
  - the sample of same-author "non-fake" articles has an average authenticity score of 33.
- Based on the training sample, the authors set cutoffs:
  - $\Pr(F) < 0.01 \Rightarrow$ "Non-fake"
  - $0.01 \leq \Pr(F) \leq 0.20 \Rightarrow$ "Ambiguous"
  - $\Pr(F) > 0.2 \Rightarrow$ "Fake"

Test Power

		Classification		
Actual	(total)	F	Ambig.	NF
Fake	171	17	137	17
Non-Fake	334	1	185	148
Total		18	322	165

- For the training-sample, the size (Pr(Type II)) is small:
  - 1/334 Non-Fake articles are classified as Fake.
- However, consistent with the 20% cutoff, a lot of the fake articles are classified as non-fake.
- For the test-sample articles, 2.8% are classified as Fake.
  - *Question:* given the 2.8% unconditional prob., can we be sure that 0% of the "non-fake" training sample really are NF?
  - Do the fake-classified articles cluster by author?
    - Jayson example suggests they should (2/31,000 classified fake)

Test Power

		Classification		
Actual	(total)	F	Ambig.	NF
Fake	171	17	137	17
Non-Fake	334	1	185	148
Total		18	322	165

- For the training-sample, the size (Pr(Type II)) is small:
  - 1/334 Non-Fake articles are classified as Fake.
- However, consistent with the 20% cutoff, a lot of the fake articles are classified as non-fake.
- For the test-sample articles, 2.8% are classified as Fake.
  - *Question:* given the 2.8% unconditional prob., can we be sure that 0% of the "non-fake" training sample really are NF?
  - Do the fake-classified articles cluster by author?
    - Jayson example suggests they should (2/31,000 classified fake)

Results

- Volume and volatility more increase significantly on days $t - t{+}2$ relative to a SA or MF article release date.
  - The increase is far higher for small firms, where volume & volatility double.
- Increase in volume is somewhat higher for LIWC-fake articles, on small firms with more retail investors.
- Increase in volatility is also slightly higher for fake articles
  - ...for small firms with more retail investors.
  - However, this is not statistically significant.
- Following SEC lawsuit, article reaction (volume & volatility) become much more muted.

Results

- Volume and volatility more increase significantly on days $t - t{+}2$ relative to a SA or MF article release date.
  - The increase is far higher for small firms, where volume & volatility double.
- Increase in volume is somewhat higher for LIWC-fake articles, on small firms with more retail investors.
- Increase in volatility is also slightly higher for fake articles
  - ...for small firms with more retail investors.
  - However, this is not statistically significant.
- Following SEC lawsuit, article reaction (volume & volatility) become much more muted.

Why Volatility?

- The authors examine volatility around the publication of the LIWC-fake stories.
- The authors argue:

  The dependent variable is the sum of daily idiosyncratic volatility on the day the article is published plus the next two days. This analysis captures whether articles moved prices around the days they were published. We examine price volatility as opposed to returns because it is exceedingly difficult to sign the direction of the content of the articles. (p.23)

- However, a quick scan of the article titles suggests that almost all of these stories are positive
  - This is also what would be anticipated under the authors' hypothesis about the nature of the fake-news

Why Volatility?

- The authors examine volatility around the publication of the LIWC-fake stories.
- The authors argue:

  > *The dependent variable is the sum of daily idiosyncratic volatility on the day the article is published plus the next two days. This analysis captures whether articles moved prices around the days they were published. We examine price volatility as opposed to returns because it is exceedingly difficult to sign the direction of the content of the articles. (p.23)*

- However, a quick scan of the article titles suggests that almost all of these stories are positive
  - This is also what would be anticipated under the authors' hypothesis about the nature of the fake-news

Fake-news story titles

First eight story titles:

- "Ampio Pharmaceuticals' Bevy of Repositioned Products Could Make a Splash in Health Care"
- "Biotech-on-The-Verge Ampio Pharmaceuticals Bolsters Its Management"
- "Ampio Rebounds As Study Results Published in Europe"
- "Ampio Pharmaceuticals: Catalysts/Recent Slide May Have Opened Up A Buying Opportunity"
- "Ampio's Rebound Goes Hand-In-Hand With Pipeline Progress"
- "Another Round of Milestone News Solidifies Ampio's Pipeline Potential"
- "ImmunoCellular Therapeutics' Cancer Stem Cell Drug Garnering Much Attention"
- "3 Cancer Immunotherapy Biotechs with Catalysts for 2012"

- Based on a quick scan of the 494 story titles in the appendix, none are negative.
- So it seems likely that the power would be higher were you to look at r, not r^2.
  - also, separately estimating the changes in σ_r^2 and μ_r might provide a more hints as to what is actually driving these patterns.

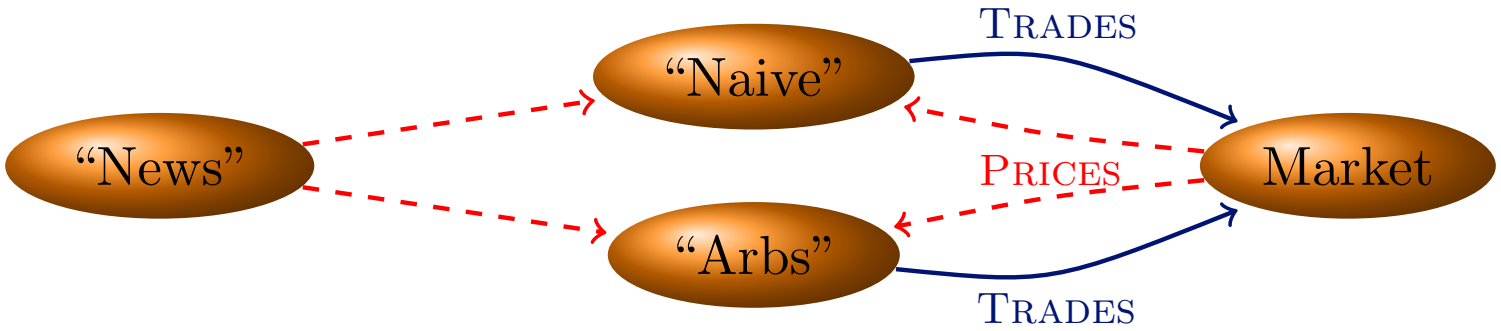
The Economic Story

- Full RE, representative agent world.
- In this case we should see:
  - depending on whether the "fake" news contains cash-flow relevant information ...

	fake news generates extra ...		
	volume	volatility	mean-reversion
no-CF-info	N	N	N
CF-info	N	Y	N

- i.e., inconsistent with the robust additional abnormal volume associated with fake-news.

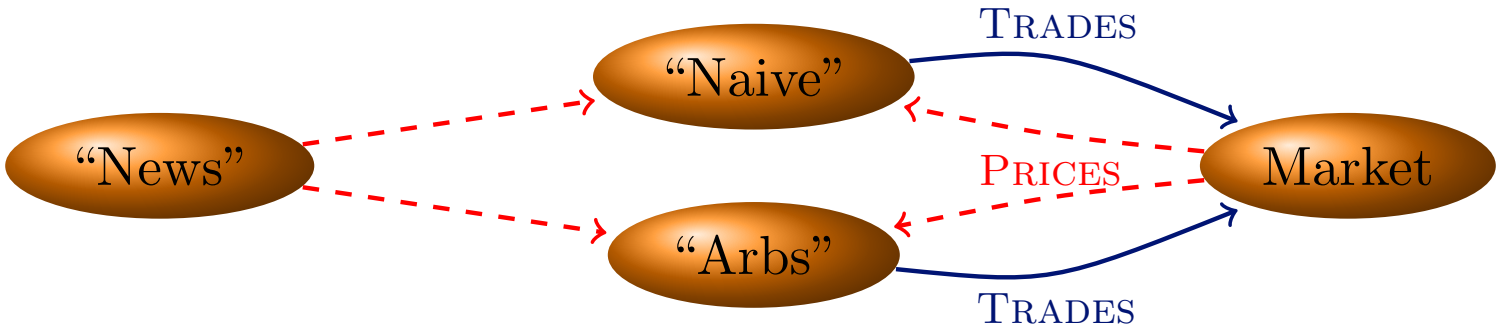
The Economic Story



Suppose we add some disagreement:

- risk-averse "Arbs" can figure out when news is fake, "Naive" investors can't.
  - Naive agents also can't extract info from prices.
- Naive agents and Arbitrageurs trade against each other.

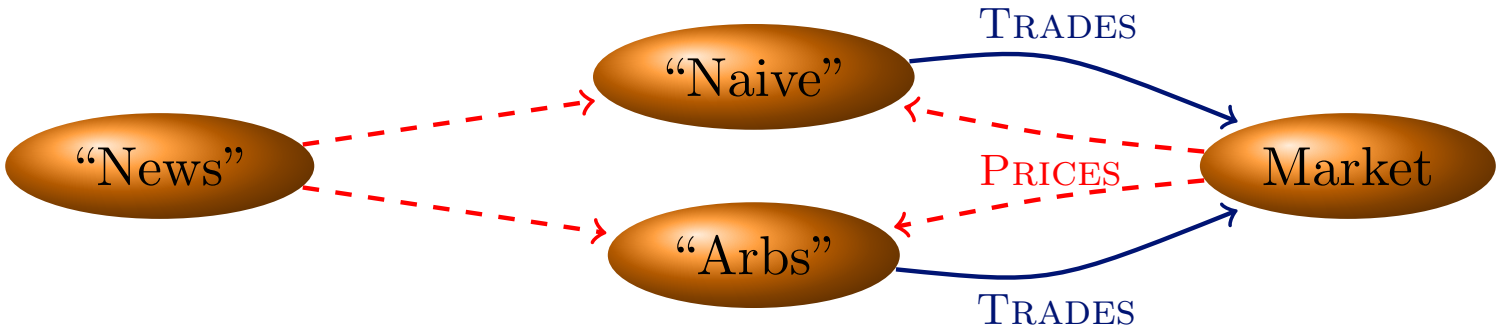
The Economic Story



- If the Arbs dominate in the economy, then:

	fake news generates extra ...		
	volume	volatility	mean-reversion
no-CF-info	Y	N	N
CF-info	Y	Y	N

The Economic Story



- If the Arbs dominate in the economy, then:

	fake news generates extra ...		
	volume	volatility	mean-reversion
no-CF-info	Y	N	N
CF-info	Y	Y	N

- If, however, the naive investors have sufficient capital:

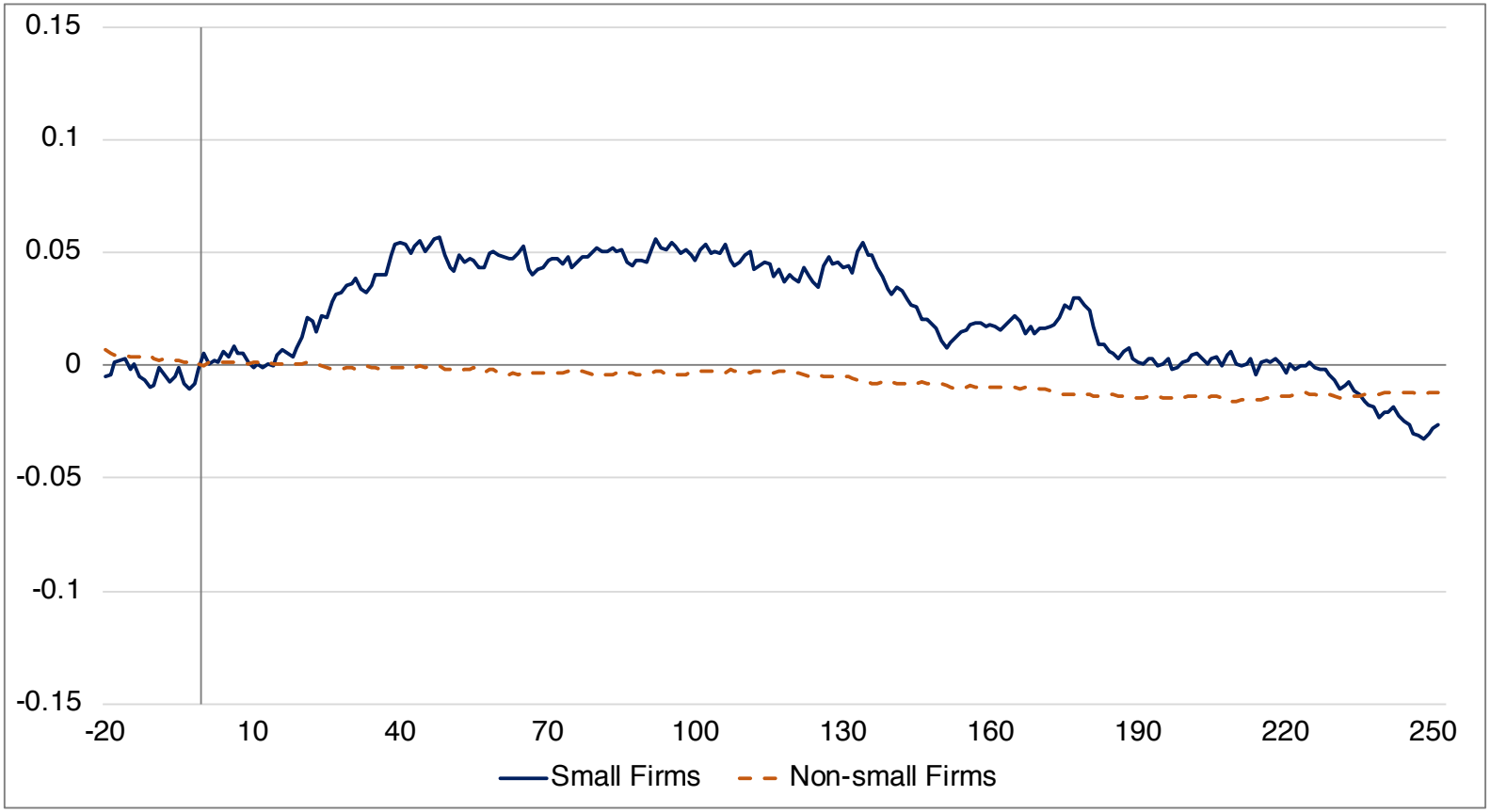
	fake news generates extra ...		
	volume	volatility	mean-reversion
no-CF-info	Y	?	Y
CF-info	Y	Y	Y

Temporary Price Moves – for-sure Fake news

Panel A: For-Sure Fake Articles from the SEC

Temporary Price Moves – LIWC Fake news

Panel B: Probabilistically Fake Articles from LIWC Algorithm



Equilibrium vs. Evolution

- As financial-economics scholars, we generally tend to think in terms of equilibrium.
  - We spend less time thinking about how we get to the equilibrium.
- This is a fascinating period of time and dataset precisely because it allows us to see how the market changes in response to the availability of new data and new techniques to analyze these data.
- Chen, De, Hu, and Hwang (2014) find that, from 2005-2012 trading on SA stories generated large abnormal returns.
  - presumably reflecting underreaction to SA-content in this period.
- Presumably, the market learned the value of this information over time ...
  - leading to a larger reaction to SA articles/comments(?), and a decline in predictability (?)
  - and perhaps leading to the use of the SA platform for fake-news.
- As documented here, this then led to a more muted reaction to the SA content (less predictability?)

Other related studies

- Mitts (2019) — "Short and Distort"

  I show how pseudonymity undermines reputational accountability in financial markets. I examine 2,900 attack articles against mid- and large-cap firms published on a website, Seeking Alpha, and show that pseudonymous ones are followed by stock-price declines and sharp reversals.

Other related studies

- Kim (2019) — "The Investment Value of Spamming Consumer Opinions"

  Using nearly 45 million reviews from Amazon.com, I measure the likelihood that a review is spam via machine learning . . . a portfolio that goes long on stocks with high abnormal review scores and low spamicity and short on stocks with low abnormal review scores and high spamicity earns abnormal returns of 1.17% to 1.23% per month.

Other related studies

- Glasserman and Mamaysky (2017) – "Does unusual news forecast market stress?" — entropy measure.

  An increase in "unusual" news with negative sentiment predicts an increase in stock market volatility.

- Li (2008) — "Fog Index"

  I find that the annual reports of firms with lower earnings are harder to read (i.e., they have higher Fog and are longer). Moreover, the positive earnings of firms with annual reports that are easier to read are more persistent. This suggests that managers may be opportunistically choosing the readability of annual reports to hide adverse information from investors.

Does the market understand the fake-news classification?

- Figure 5 A and B
- A is not an implementable strategy.
- B is, sort of!!
  - Is B just picking up salience effect (eg., Frazzini and Lamont (2007))

References I

Cecchini, Mark, Haldun Aytug, Gary J Koehler, and Praveen Pathak, 2010, Making words work: Using financial text as a predictor of financial events, *Decision Support Systems* 50, 164–175.

Chen, Hailiang, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media, *The Review of Financial Studies* 27, 1367–1403.

Frazzini, Andrea, and Owen A Lamont, 2007, The earnings announcement premium and trading volume, NBER working paper #13090.

Glasserman, Paul, and Harry Mamaysky, 2017, Does unusual news forecast market stress?, *Journal of Financial and Quantitative Analysis* pp. 1–38.

Kim, Dawoon, 2019, Not all investors know: The investment value of spamming consumer opinions, Cornell University Working Paper.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and economics* 45, 221–247.

Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.

Mitts, Joshua, 2019, Short and distort, Columbia Law and Economics Working Paper.

Pennebaker, James W, Roger J Booth, and Martha E Francis, 2007, Linguistic inquiry and word count: LIWC 2007, University of Texas Austin working paper, retrieved from http://www.depts.ttu.edu/psy/lusi/files/LIWCmanual.pdf.