



The Information Content of Corporate Websites

Bradford Lynch-Levy and Daniel Taylor

NBER Big Data, December 10, 2021

- Securities and Exchange Commission
(Release 33-8128, 2002)

“

[A]n efficient and economical method for companies to make information available about themselves to many investors is through their Internet websites.

”

Motivation

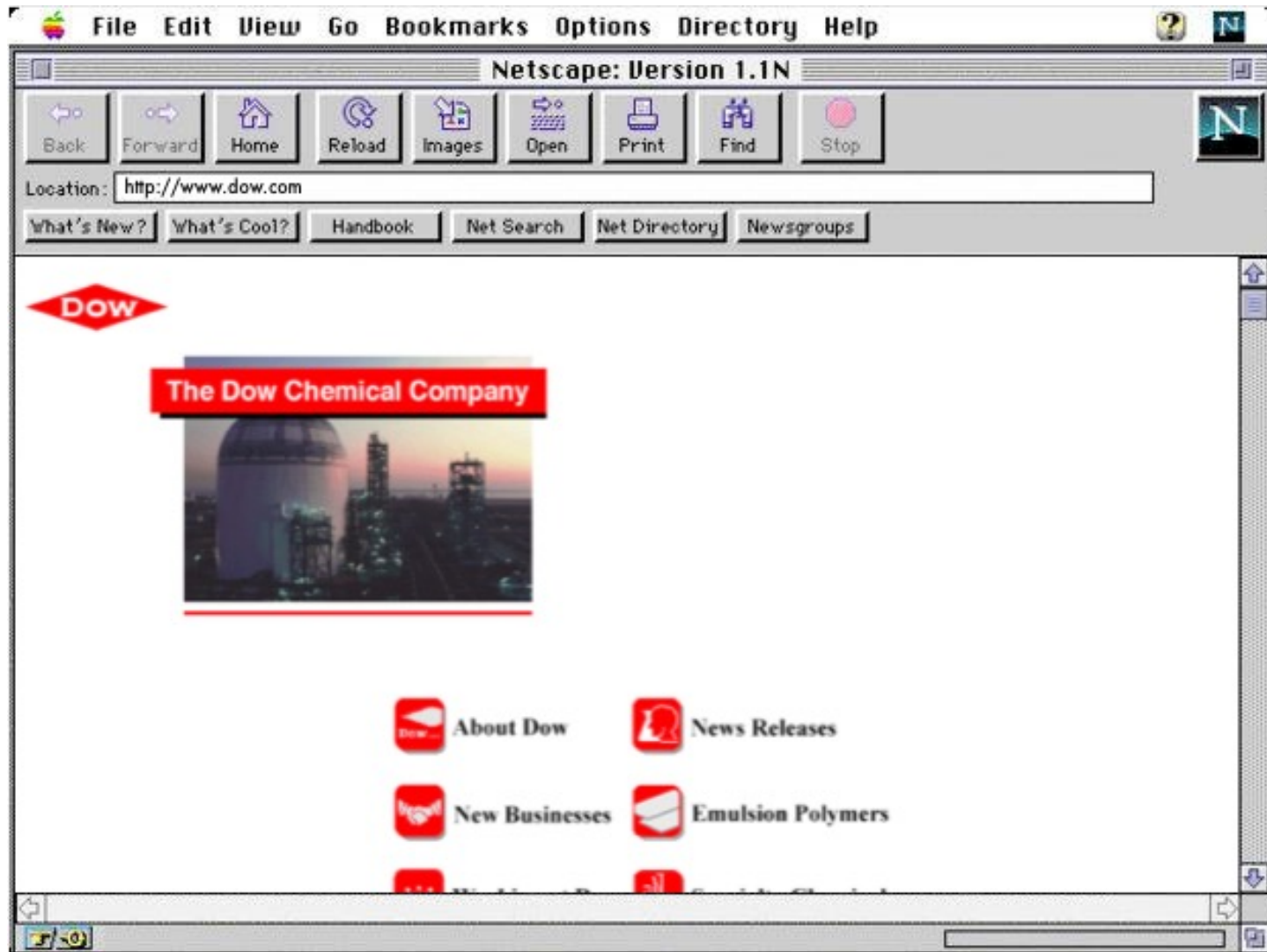
- The disclosure literature is vast
- The information effects of SEC filings, management forecasts, analyst reports, and press releases are well documented
- A survey of the literature reveals that comparatively little is known about how firms and shareholders use corporate websites

Background - Regulation

- In 1995, the SEC highlighted the growing prevalence of the PC in households and the low cost of electronic communication^{*}
- In 2000, the SEC recommends all firms use websites for investor relations^{**}
- In 2002, the SEC asks all firms to
 1. Disclose their web address in their annual report
 2. Firms without a website must explain why they do not have one^{***}
- As of August 7, 2008, the SEC allows companies to substitute website disclosures for 8-K filings on EDGAR^{****}

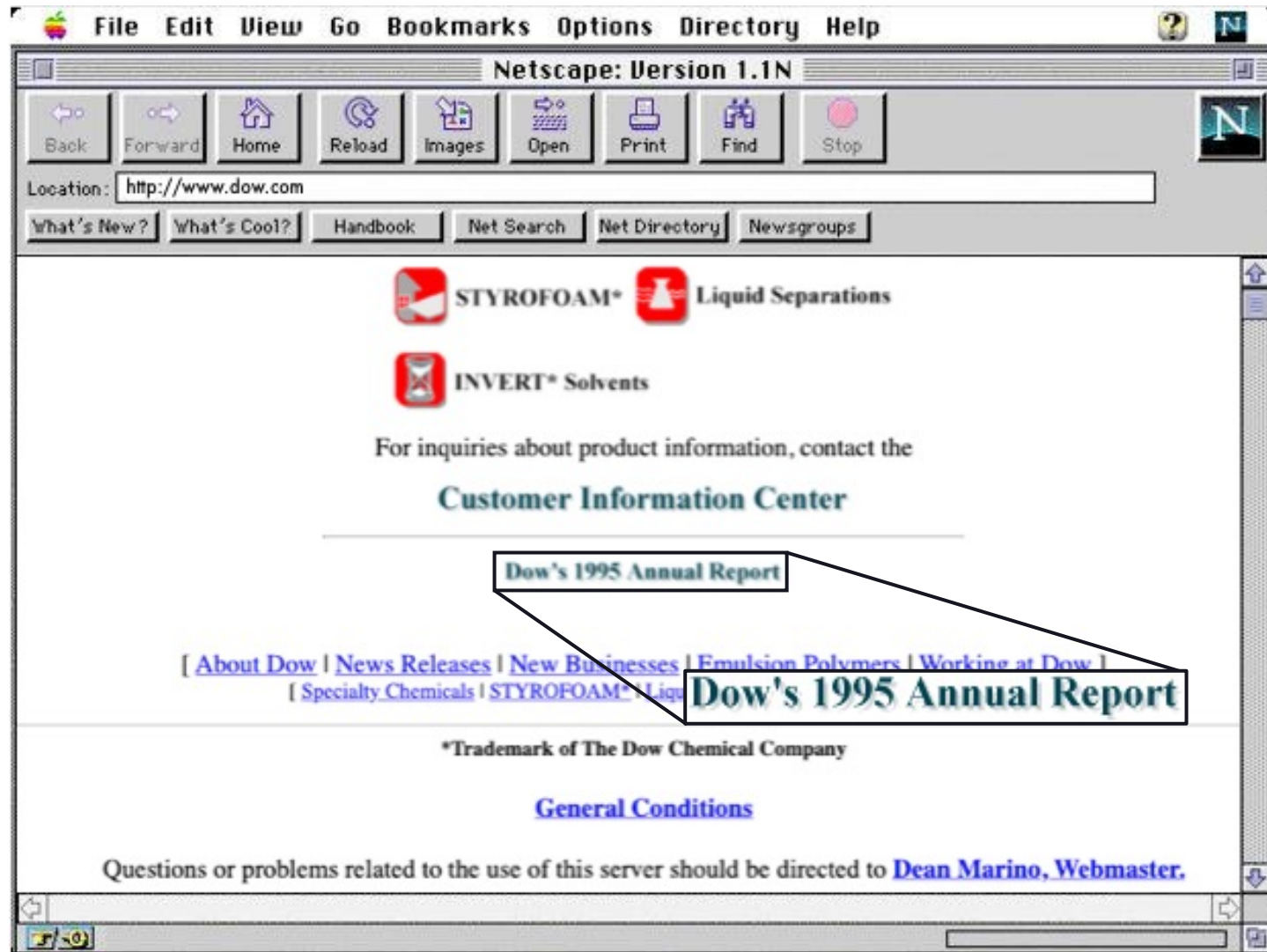
^{*}SEC Release 33-7233, ^{**}SEC Release 33-7856, ^{***}SEC Release 33-8128, ^{****}SEC Release 34-58288

Background – Adoption of websites



Dow's website circa 1996, archived by Alexa Internet.

Background – Adoption of websites



Dow's website circa 1996, archived by Alexa Internet.

But are websites more than EDGAR?

“This **code of ethics** and the Company’s **corporate governance policies** are posted on the Company’s website at <http://www.JohnDeere.com>.” (Deere & Co 2003 10-K)



“We make our website content available for **information purposes only**. It should not be relied upon for investment purposes, **nor is it incorporated by reference into this Form 10-K.**”
(Telos 10-K, 2003)

“In addition, the Company has posted the **charters** for its **Audit Committee, Compensation Committee, and Nominating and Corporate Governance Committee**, as well as the ... on its website.” (Ansys’s 2018 10-K)



But are websites more than EDGAR? - Kiora

- Kiora (Formerly EyeGate) never posts their earnings announcements to EDGAR. Instead, they keep records of the PRs on their website



ABOUT TECHNOLOGY ▼ PIPELINE INVESTORS ▼ NEWS

Kiora Pharmaceuticals Reports Third Quarter 2021 Financial Results and Provides Business Update

SALT LAKE CITY, UT, Nov 15, 2021 – Kiora Pharmaceuticals, Inc. (NASDAQ: KPRX), ("Kiora"), announced today its financial results for the quarter ended September 30, 2021 and provided an update on research and development progress.

"We recently sharpened the focus of our pipeline strategy, which now includes development of a potentially expedited and cost-efficient path to market," said Brian M. Strem, CEO of Kiora. "We anticipate multiple upcoming milestones, including advancing KIO-301 to Phase 1/2a proof-of-concept study in September to evaluate surface inflammation."

Recently, Kiora has achieved several milestones, including the following:

Pipeline Updates

- Acquired Bayon Therapeutics, Inc. in October, which brought in a portfolio of small molecule drugs, which is designed to restore vision in patients with inherited and age-related degenerative diseases.
- Completed enrollment in a Phase 1/2a proof-of-concept study in September to evaluate surface inflammation.

Corporate Updates

- Appointed Eric J. Daniels, M.D. as Chief Development Officer in October to drive the portfolio of clinical-stage assets, expanding pipeline, and proprietary platform.
- Raised gross proceeds of \$10.75 million through a registered direct offering in August.
- Appointed Brian M. Strem, Ph.D. as CEO in July bringing to the Company strategic and scientific expertise in ophthalmology.



EDGAR Search Results

Home | Latest Filings | Previous Page

U.S. Securities and Exchange Commission

Give us feedback on EDGAR [CLICK HERE](#)

[SEC Home](#) » [Search the Next-Generation EDGAR System](#) » [Company Search](#) » [Current Page](#)

KIORA PHARMACEUTICALS INC CIK#: 0001372514 (see all company filings)

SIC: 2834 - PHARMACEUTICAL PREPARATIONS
State location: UT | State of Inc.: DE | Fiscal Year End: 1231
formerly: EYEGATE PHARMACEUTICALS INC (filings through 2021-11-04)
(Office of Life Sciences)
[Get insider transactions](#) for this issuer.

Business Address
1371 EAST 2100 SOUTH
SUITE 200
SALT LAKE CITY UT 84105
(781) 788-8869

Mailing Address
1371 EAST 2100 SOUTH
SUITE 200
SALT LAKE CITY UT 84105

Filter Results

Filing Type: 8-K
Prior to: (YYYYMMDD) 20211130

Ownership?
☐ include
☒ exclude
only

Limit Results Per Page
20 Entries

[Search](#)
[Show All](#)

Search Within Files

[EDGAR](#) | [Full Text Search](#)

Enter keywords

[Search](#)

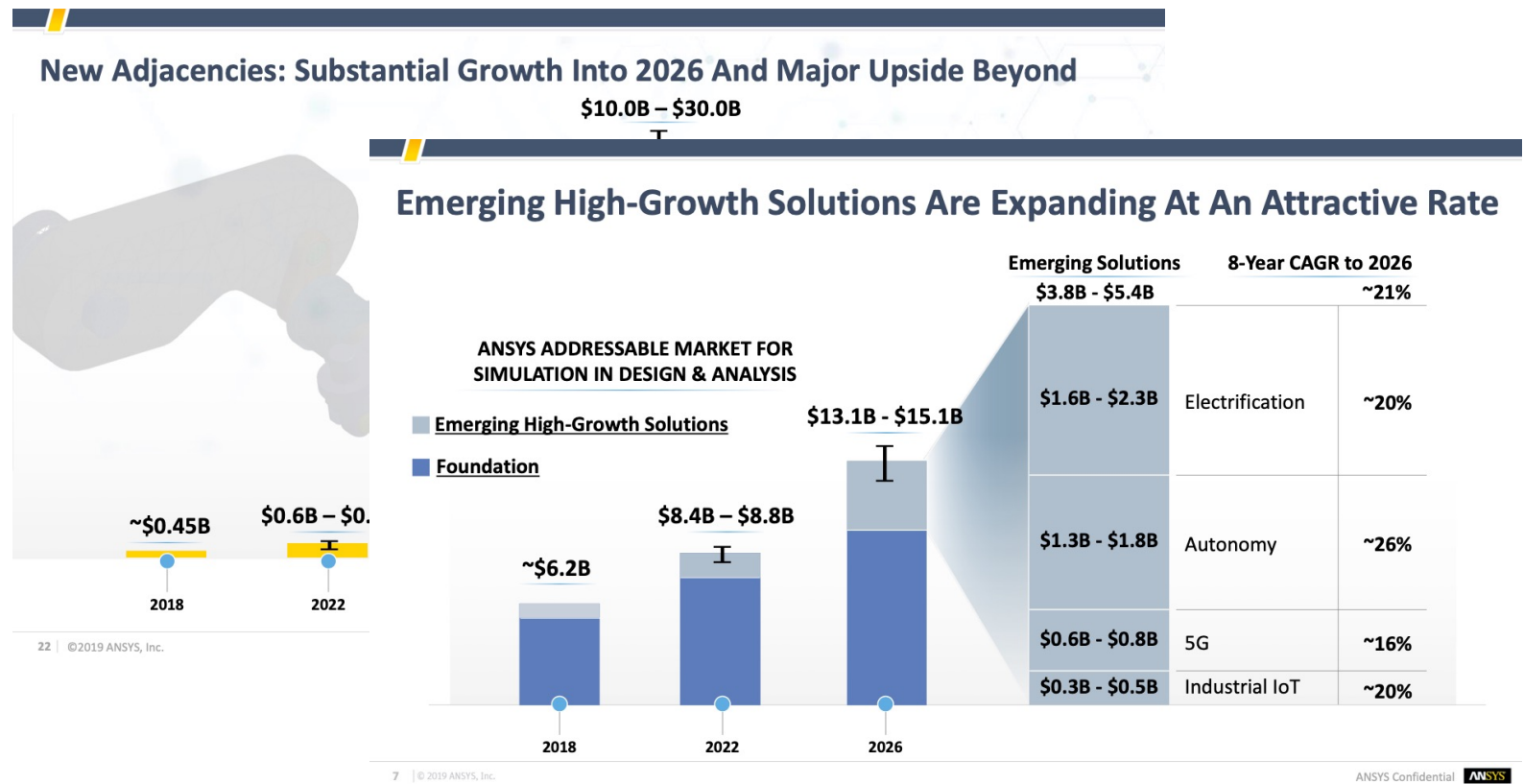
Items 1 - 20 [RSS Feed](#)

[Next 20](#)

Filings	Format	Description	Filing Date	File/Film Number
8-K	Documents Interactive Data	Current report, items 5.03, 8.01, and 9.01 Acc-no: 0001104659-21-135333 (34 Act) Size: 430 KB	2021-11-08	001-36672 211386528
8-K	Documents Interactive Data	Current report, items 1.01, 2.01, 3.02, 5.02, 7.01, and 9.01 Acc-no: 0001104659-21-129738 (34 Act) Size: 677 KB	2021-10-26	001-36672 211345899
8-K	Documents Interactive Data	Current report, items 7.01 and 9.01 Acc-no: 0001104659-21-115504 (34 Act) Size: 7 MB	2021-09-14	001-36672 211252608

But are websites more than EDGAR? - Ansys

- 10+ presentations containing hundreds of slides by top management were posted to Ansys's website and not mentioned elsewhere



Research Questions

- Do firms use their corporate website to provide pertinent new information to investors?
- If so, what are the broad topics of that information?

Contribution

1. Investor relations

- Corporate website disclosure is economically significant
- Websites contain *more* business operations content than EDGAR

2. Information processing costs

- Websites alleviate information asymmetry between investors

3. Effects of new technologies on markets

- Websites are largely unregulated but have significant capital market effects

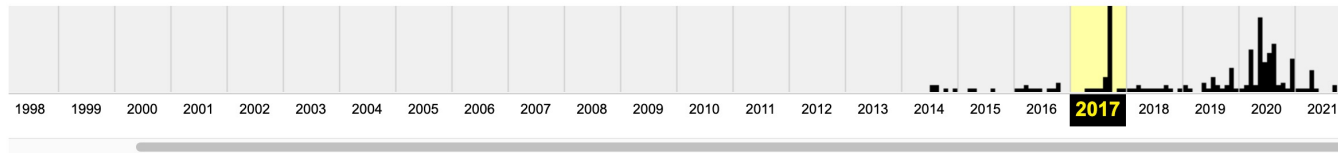
These results hold across a variety of research designs, identification strategies, and cross-sectional tests

Related Literature on Corporate Websites

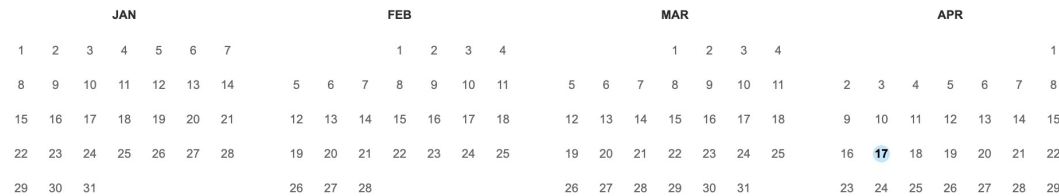
- Boulland et al. (working paper) construct a **quarterly** measure of website content using data from the Internet Archive (IA)
 - Find it is positively correlated with the provision of earnings forecasts, EDGAR filings, and EDGAR file size
- The IA, or Wayback Machine, scrapes webpages and makes historical archives available to the public
- Caution: IA endogenously chooses which pages to scrape and when to scrape them based on internet attention
 - IA data poses key identification challenges

Related Literature – *Celcuity Inc.*

Saved **214 times** between [July 24, 2014](#) and [November 27, 2021](#).

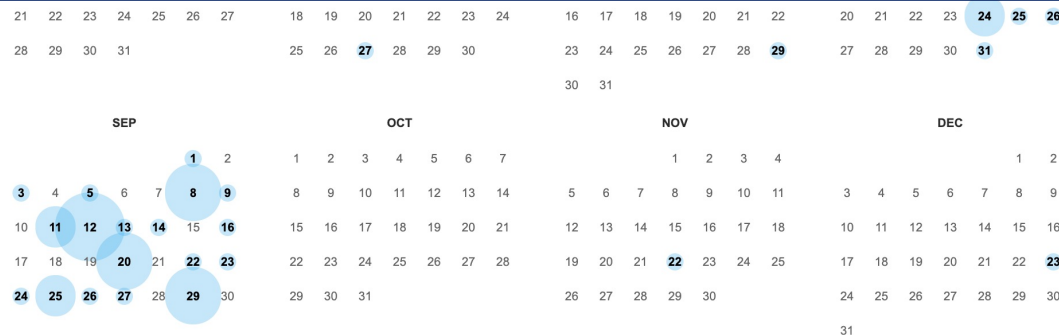


Wed, 22 Nov 2017 03:03:18 GMT (why: No Collection Info)



Note

This calendar view maps the number of times **celcuity.com** was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the [FAQ](#).

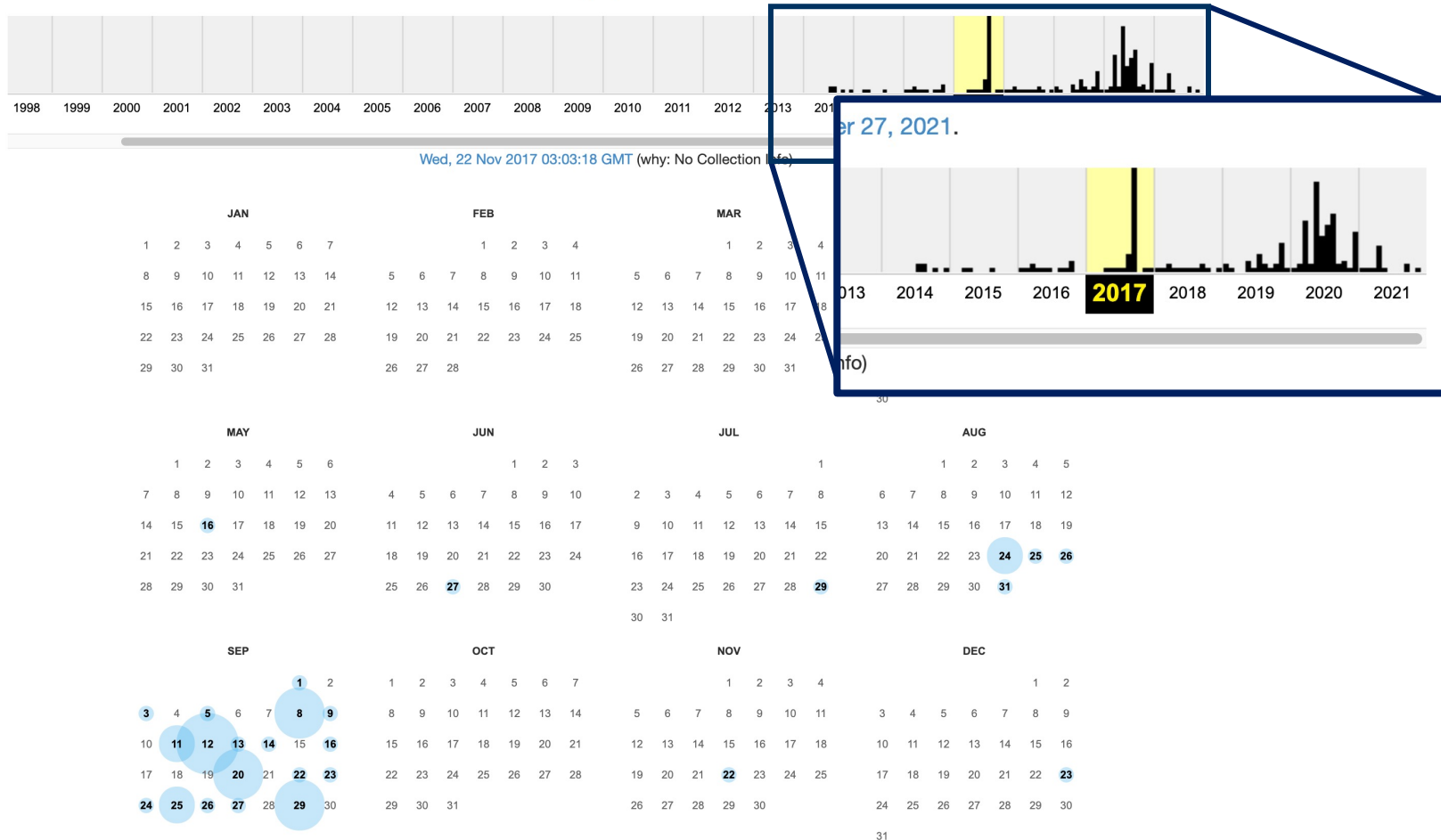


Note

This calendar view maps the number of times **celcuity.com** was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the [FAQ](#).

Related Literature – *Celcuity Inc.*

Saved 214 times between July 24, 2014 and November 27, 2021.

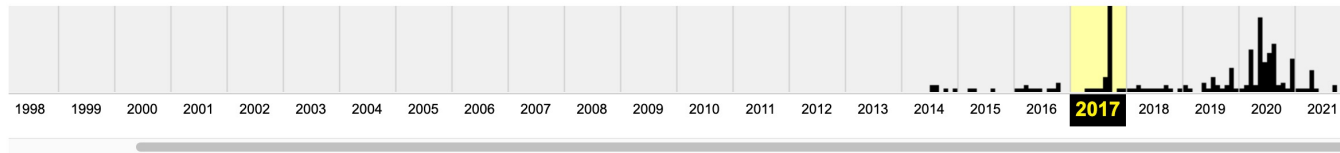


Note

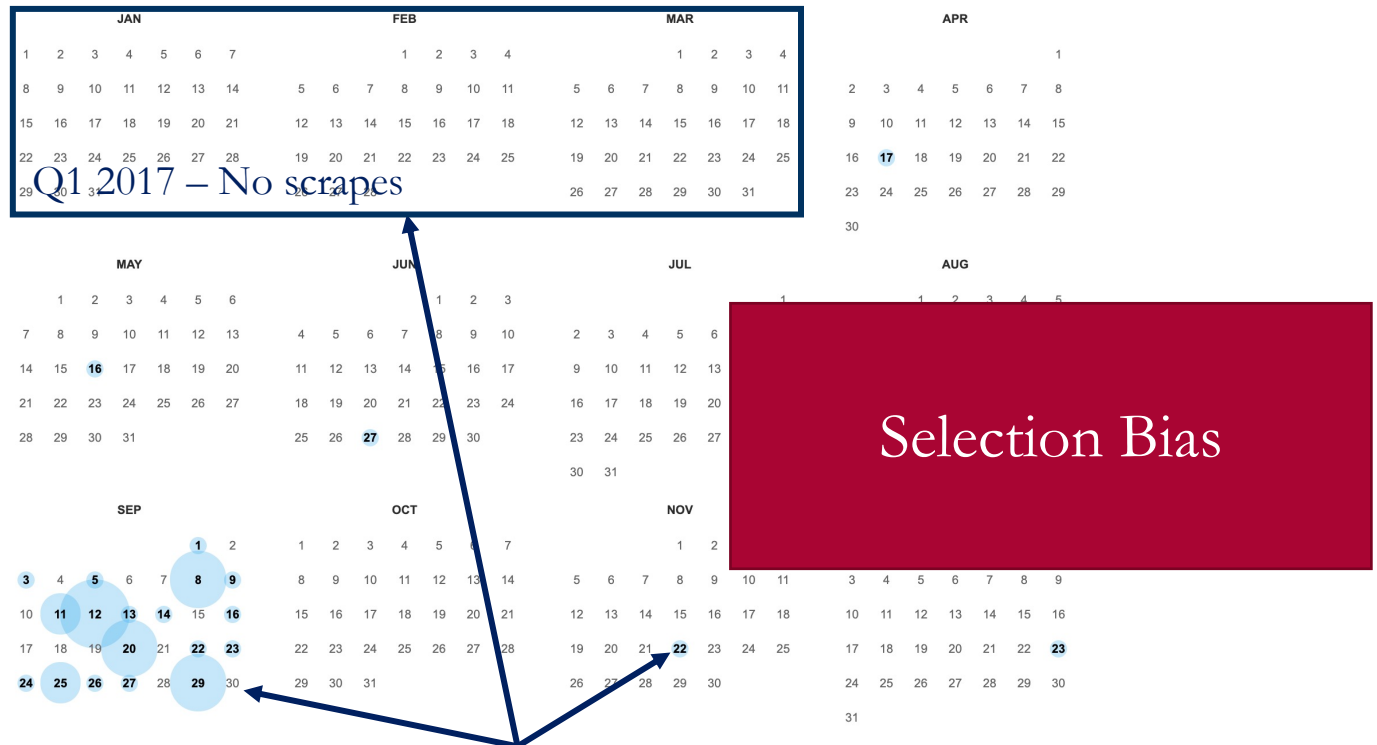
This calendar view maps the number of times **celcuity.com** was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the [FAQ](#).

Related Literature – *Celcuity Inc.*

Saved 214 times between July 24, 2014 and November 27, 2021.



Wed, 22 Nov 2017 03:03:18 GMT (why: No Collection Info)



Note

This calendar view maps the count of scrapes that day.

Bubble size represents
count of scrapes that day

of how many times the

Sample Selection

- We collect daily observations of *all* corporate websites listed in Compustat *everyday* at 9:15AM from July through October, 2019
 - Minimal selection bias
 - Includes the homepage and all inner pages
- Sample is ~150 Terabytes
- We collect data on control variables from Compustat, CRSP, RavenPack, SEC EDGAR, and NYSE TAQ
- 220,505 firm-date observations

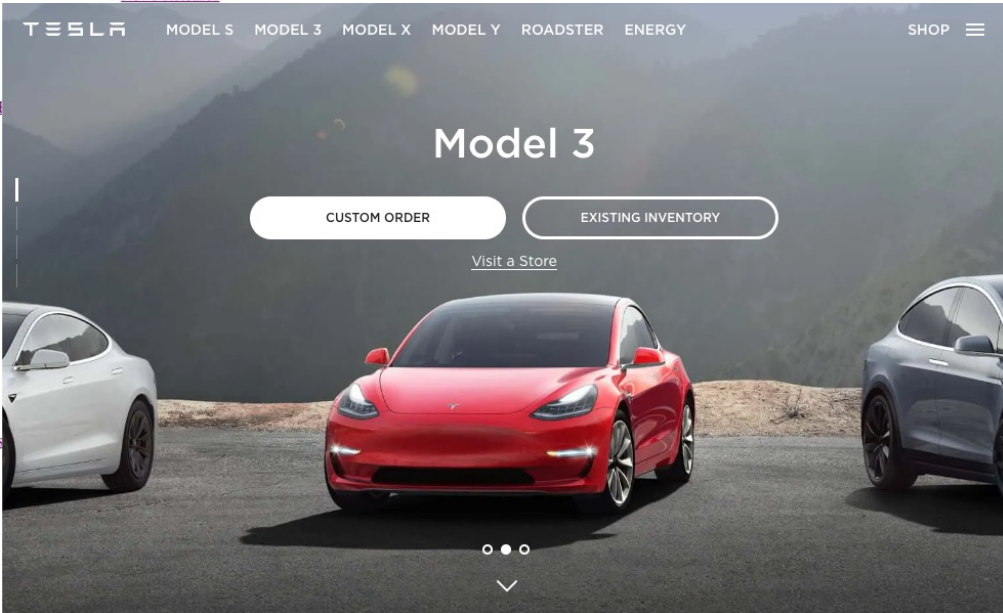
Measuring Website Content

- In principle, there are numerous ways to measure content
 - Text, images, videos, tables, external data referenced via hyperlinks

BERKSHIRE HATHAWAY INC.
3555 Farnam Street
Omaha, NE 68131
Official Home Page

- [A Message From Warren E. Buffett](#)
- [Annual & Interim Reports](#)
Updated August 3, 2019
- [Special Letters From Warren & Charlie RE: Past, Present & Future](#)
- [Link to SEC Filings](#)
- [Links to Berkshire Subsidiary Companies](#)
- [Corporate Governance](#)
- [Sustainability](#)
- [Owner's Manual](#)
- [Letters from Warren E. Buffett Regarding Pledges of Gifts of Berkshire Stock](#)

- [News Releases](#)



The image shows a screenshot of the Tesla website's Model 3 page. The header features the Tesla logo and navigation links for Model S, Model 3, Model X, Model Y, Roadster, Energy, and Shop. The main heading is 'Model 3'. Below it are two buttons: 'CUSTOM ORDER' and 'EXISTING INVENTORY'. A link 'Visit a Store' is positioned below the buttons. The background image shows three Tesla Model 3 cars (white, red, and blue) parked on a road with mountains in the background. At the bottom, there are three small circles and a downward arrow.

Measuring Website Content

- In principle, there are numerous ways to measure content
 - Text, images, videos, tables, external data referenced via hyperlinks
- We measure the length of the document, the number of HTML formatting tags, links, rich content, and dynamic elements

Variable	N-obs	Mean	Std	Median
<i>Length</i>	220,505	2791.14	4335.93	1773
<i>Formatting</i>	220,505	372.84	381.47	281
<i>Links</i>	220,505	123.59	144.95	89
<i>Rich</i>	220,505	32.92	71.84	16
<i>Dynamic</i>	220,505	20.44	18.72	16

Measuring Website Content (Cont'd)

- We construct a composite measure, *WebContent*, using factor analysis
- The first principal factor explains 92% of the variation in our five proxies for content, loads positively on all measures, and is the only factor with an eigenvalue greater than unity

Factor	Eigenvalue	Proportion of the variation explained	Cumulative proportion of the variation explained	First Factor Loading	
				Variable	Weight
1 st	2.28	92.4%	92.4%	<i>Length</i>	0.527
2 nd	0.16	6.6%	99.0%	<i>Formatting</i>	0.948
3 rd	0.03	1.0%	100.0%	<i>Links</i>	0.845
				<i>Rich</i>	0.471
				<i>Dynamic</i>	0.833

Daily Event Study

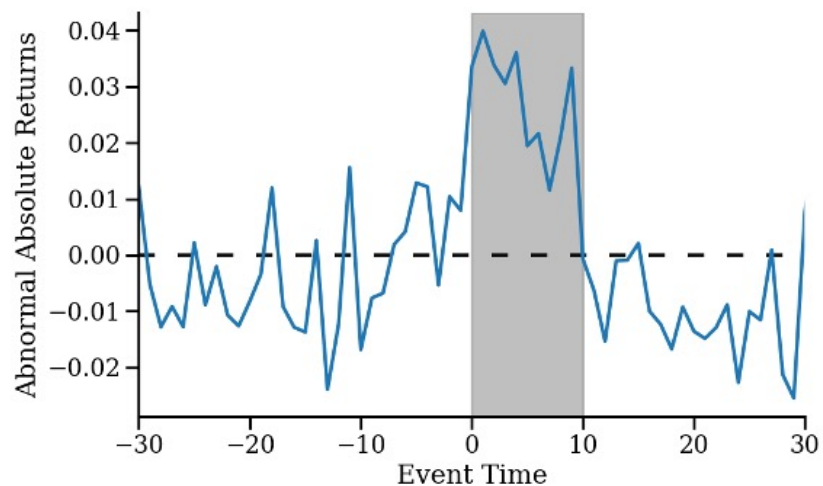
- Focus on “events” where the change in *WebContent* is in the top tercile and there are no EDGAR filings or PRs in the $[-1,+1]$ window around the change
- Examine changes in capital market outcomes in a tight window around the event
- Daily nature of the data allows for firm-quarter fixed effects
 - Tight event study design to test whether websites contain information content incremental to SEC filings and PRs

Results

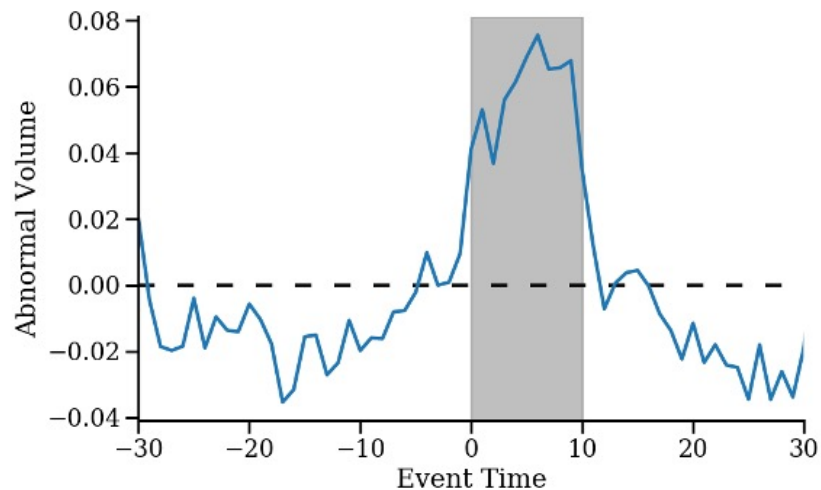
$$Outcome_{i,t+1} = \alpha + \gamma \cdot Controls_{i,t} + \varepsilon_{i,t}$$

$$Plot_t = \frac{E[\varepsilon_t] - E[\varepsilon]}{Var(\varepsilon)}$$

Absolute Returns



Trading Volume

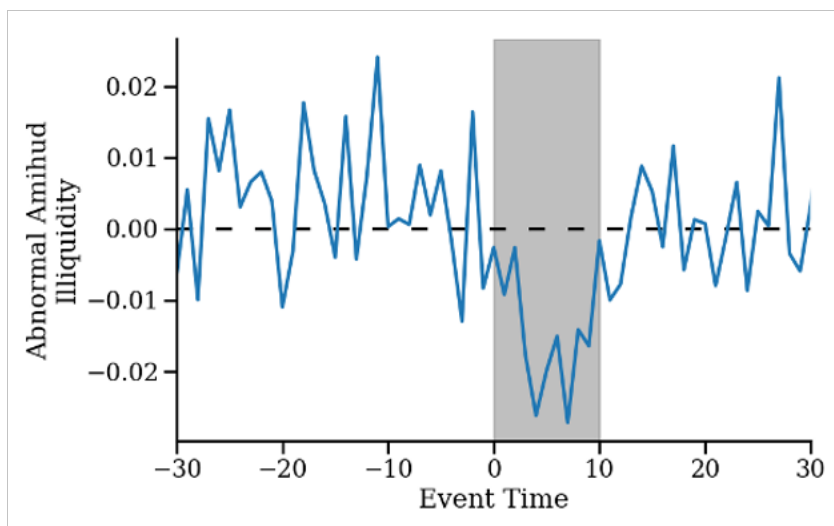


Results

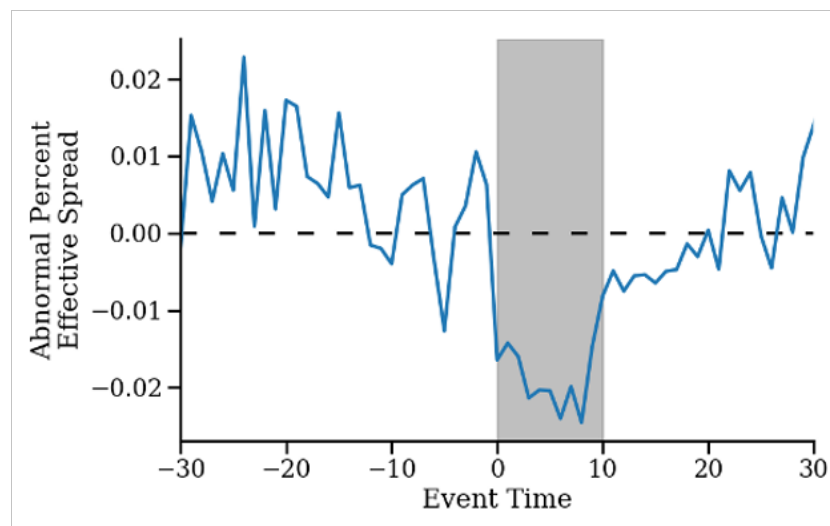
$$Outcome_{i,t+1} = \alpha + \gamma \cdot Controls_{i,t} + \varepsilon_{i,t}$$

$$Plot_t = \frac{E[\varepsilon_t] - E[\varepsilon]}{Var(\varepsilon)}$$

Amihud Illiquidity

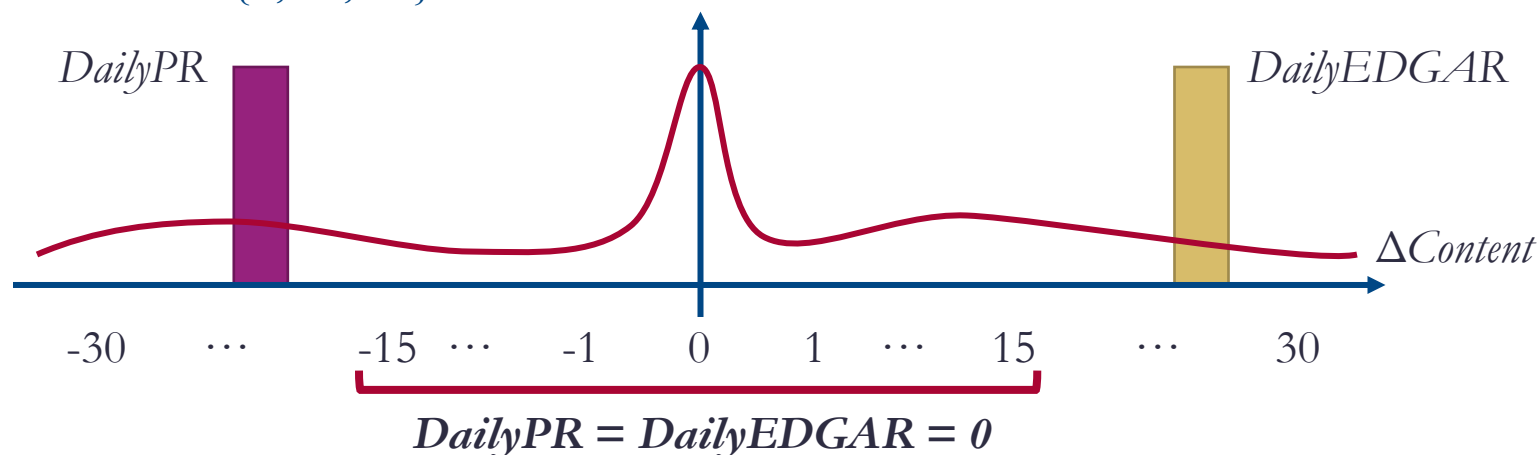


Effective Spread



Dissemination Versus New Information

- It is possible that corporate websites merely serve as a dissemination tool rather than providing new information
- To distinguish between the two roles, we re-estimate our previous models while varying the width of the non-disclosure window
 - Require that there are no SEC filings or PRs in the $[-s, +s]$ window around the change in website content
 - Where $s \in \{5, 10, 15\}$

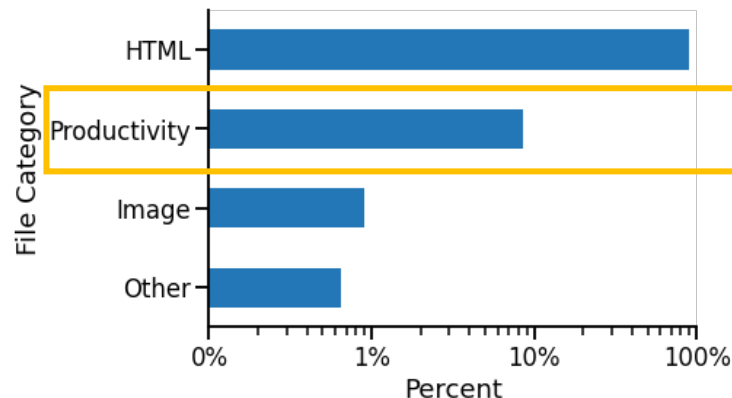


Dissemination Versus New Information

	Dep. Var. <i>AbsoluteReturn</i>		Dep. Var. <i>Volume</i>	
	Coeff. on	<i>t</i> -stat on	Coeff. on	<i>t</i> -stat on
<i>Non-disclosure Window [-s,+s]</i>	<i>EventDay[0,10]</i>	<i>EventDay[0,10]</i>	<i>EventDay[0,10]</i>	<i>EventDay[0,10]</i>
Excluding observations with disclosures [-5,+5] <i>N-obs</i> = 444,446, <i>N-events</i> = 7,214	0.60***	(5.82)	0.55***	(6.51)
Excluding observations with disclosures [-10,+10] <i>N-obs</i> = 307,112, <i>N-events</i> = 4,987	0.63**	(5.00)	0.43***	(4.69)
Excluding observations with disclosures [-15,+15] <i>N-obs</i> = 228,548, <i>N-events</i> = 3,712	0.53***	(3.44)	0.30***	(2.97)
<i>Controls</i>	Yes		Yes	
<i>Date FE</i>	Yes		Yes	
<i>Firm-Year-Qtr FE</i>	Yes		Yes	

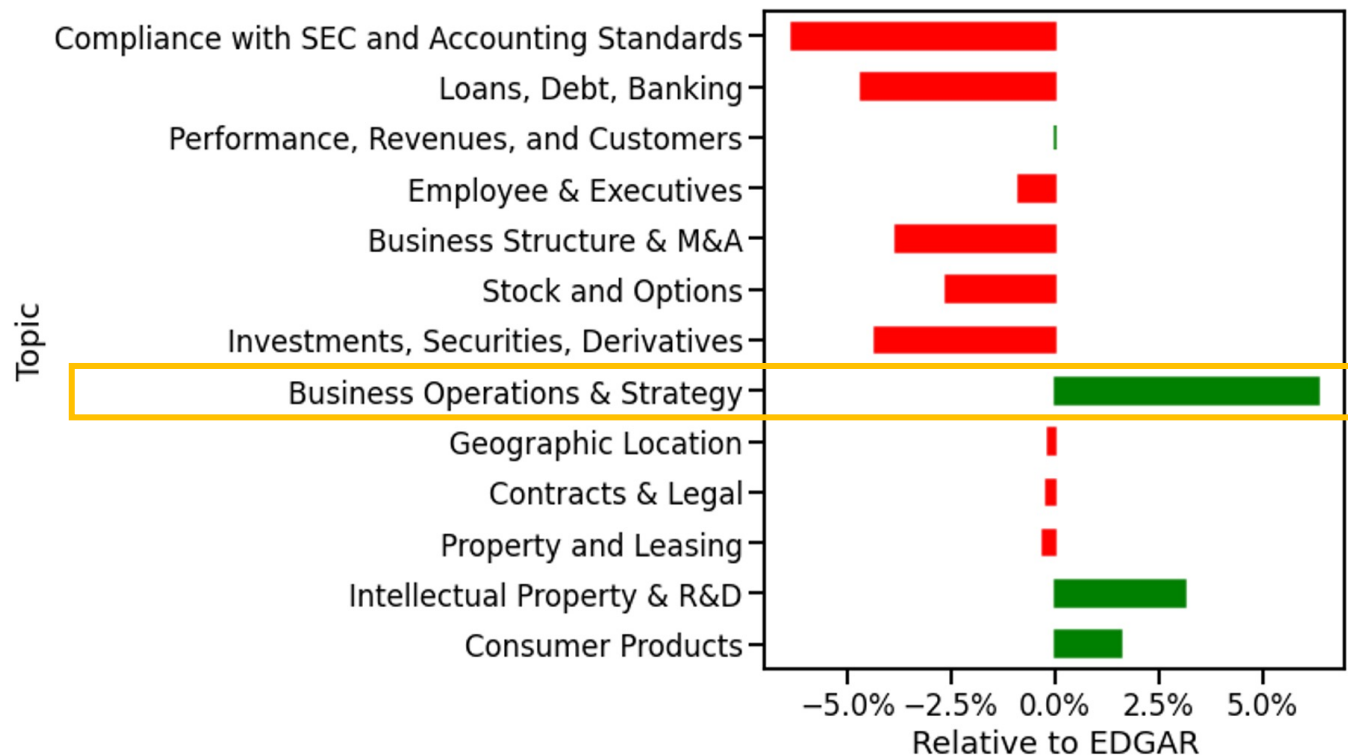
Types of Content in Corporate Websites

- We classify changes in *WebContent* according to file type, e.g., Office documents, PDFs, and engineering drawings are *Productivity* related
- *Productivity* related files make up roughly 10% of the content in *WebContent* changes



Types of Content in Corporate Websites

- We classify changes in *WebContent* using the topic model of Dyer et al. (2017) and compare to EDGAR



Conclusions

- Using a tight event study design, we find that websites contain information content incremental to traditionally studied channels
- Notably, the market reaction to website changes is of similar economic significance as SEC filings and press releases
- Additionally, websites contain *more* business operations content than EDGAR and the market reaction to such information is heightened