Corporate Websites: A New Measure of Voluntary Disclosure^{*}

Romain Boulland[†] Thomas Bourveau[‡] Matthias Breuer[§]

First version: 07/26/2019This version: 03/31/2021

Abstract

We construct a new measure of voluntary disclosure based on firms' websites. Using the Wayback Machine, we create a standardized measure of disclosure capturing the quantity of information on firms' websites. We validate our measure by documenting that it is positively associated with established measures of firms' voluntary disclosure and liquidity. Importantly, we document that our measure, while correlated with established disclosure measures, is not subsumed by those measures. It complements existing measures in three important ways. First, our measure captures not only capital-market-related but also additional information geared toward stakeholders other than investors (e.g., customers). Secondly, our measure can be calculated for a broader sample of firms, including small, private, and international firms. Lastly, our measure can be customized to fit the specific research question at hand (e.g., ESG disclosures) via textual analysis of website content.

^{*}We thank Matt Bloomfield, Ed deHaan, Henry Friedman, Mirko Heinle, Philip Joos, Christian Leuz, Bill Mayew (discussant), Anya Nakhmurina, Shiva Rajgopal, Delphine Samuels, Bryce Schonberger (discussant), George Serafeim, Christina Zhou, seminar participants at Harvard Business School, and conference participants at the 2019 Stanford Summer Camp, the 2019 Dartmouth (Tuck) Accounting Conference, and the 2019 Burton Conference at Columbia University for helpful discussions and comments. We thank Marlene Plumlee and Jing He for sharing their 8-K measures, and Shuping Chen for sharing her disclosure quality measure. The code to construct our measure is available on our GitHub repository. All errors are our own.

[†]ESSEC Business School - boulland@essec.edu

[‡]Columbia University, Graduate School of Business - tb2797@columbia.edu

[§]Columbia University, Graduate School of Business - mb4468@gsb.columbia.edu

1 Introduction

Firms have incentives to communicate information to their various stakeholders, including investors, suppliers, customers, and employees (e.g., Beyer et al., 2010; Dranove and Jin, 2010). In recent years, this communication has been facilitated by and shifted toward the internet. The internet, like other public corporate communication channels (e.g., newspapers and TV advertisements), allows firms to reach all their various stakeholders at once. Unlike other channels, it, however, allows firms to directly communicate with their stakeholders without intermediation, time lag, or restrictive constraints on the amount or format of the information disclosed. As a result, practitioners view the internet, and firms' websites in particular, as a powerful tool for firms to communicate with stakeholders (IR-Society, 2012). Although firms' websites provide a one-stop repository for firm-specific information, which tends to be the first place stakeholders look for information, the disclosure literature has not paid much attention to firms' websites as a public disclosure channel yet.

In this paper, we develop a novel measure of firms' voluntary disclosure based on firms' websites. We proceed in four steps. We first create a large database of firms' historical websites and construct a standardized website-based measure of disclosure using firms' website size. We next explore the variation underlying our website-size measure (e.g., firms' website content). We then validate our website-size measure, examining its associations with established measures of firms' voluntary disclosure (e.g., management forecasts and voluntary 8-K filings) and liquidity (i.e., bid-ask spreads) in a sample of publicly listed U.S. firms. We conclude by discussing potential applications of our website-based measure in future research.

To create a database of firms' historical websites, we use the Wayback Machine, a digital archive maintained by the Internet Archive. The Wayback Machine periodically crawls and stores snapshots of nearly all existing websites. It provides the list of all Uniform Resource Locators (URLs) constituting the website, the size (in octets or bytes) of each URL, and whether a specific URL contains text only or enriched content, such as images, videos, or specific applications (e.g. Excel files, PDF files). We obtain this information on a quarterly basis, from Q1 1997 to Q4 2018, for a sample of COMPUSTAT U.S. firms with non-missing URLs. In terms of coverage, our measure is available for approximately 34% of the CRSP-COMPUSTAT universe in 1997. This coverage improves steadily, reaching a high of 94% in 2018.

As a standardized measure of voluntary disclosure, we use the total size of firms' websites. This measure conceptually captures the quantity of information provided to stakeholders. In this regard, it is similar to existing disclosure measures capturing the length of firms' filings (e.g., Dyer et al., 2017; He and Plumlee, 2020; Breuer et al., 2020) or the number of disclosed items (e.g., Lang and Lundholm, 1993; Botosan, 1997; Chen et al., 2015). Besides our standardized measure, we can also construct two types of content-specific measures. Our first content-specific measure is based on the information contained in firms' URL strings. This measure allows us to examine broad content categories discussed on firms' websites: e.g., products, investors, human resources, and geography. Our second content-specific measure is based on textual analysis of firms' entire website content (e.g., searching the keyword 'environment' to capture ESG topics).

We start our empirical analyses by exploring the variation for our new disclosure measure. Using a sample of 214,910 firm-quarter observations, we document that our measure exhibits rich variation in size and content. To assess the sources of variation in our measure, we regress it on various combinations of industry, firm, and time fixed effects. These regressions reveal that the variation in our measure is more strongly driven by fixed firm-level factors than common time trends (e.g., a general increase in firms' online activities). This pattern is consistent with firms adopting and largely committing to a particular corporate-website disclosure strategy, in line with existing disclosure studies and theories, which tend to focus on cross-sectional differences in firms' disclosure strategies. Our measure, however, also exhibits notable within-firm variation, permitting an examination of firms' disclosure changes around salient events (e.g., using difference-in-differences designs). With respect to website content, we document that our sample firms discuss central aspects of their business on their websites. They predominantly provide product and investor related information on their websites. Other aspects (e.g., human resources), however, are also frequently discussed. Overall, there is substantial variation in the relative prominence of the distinct categories across industries. This cross-industry variation appears to be linked to industry characteristics, such as the nature of firms' assets, the geographical scale of their operations, or the importance of labor. In sum, our content exploration suggests that corporate websites contain information relevant for various stakeholders, including investors, customers, suppliers, and employees.

We validate our standardized disclosure measure by correlating it with existing disclosure measures. We find that our standardized measure is positively associated with four established measures of voluntary disclosure: the number of management earnings forecasts (Marquardt and Wiedman, 1998), the extent of disaggregation in firms' financial statements (Chen et al., 2015), the word count of voluntary 8-K filings (He and Plumlee, 2020), and the gross 10-K file size (Loughran and McDonald, 2014).¹ For all established measures except the financial-statement disaggregation measure, this positive association even holds when only focusing on within-firm changes. These results suggest that our measure captures similar information as established measures of voluntary disclosure.

We also correlate our standardized disclosure measure with bid-ask spreads as a measure of firms' share liquidity and investors' information asymmetry. Disclosure theories predict that increased public information disclosure reduces information asymmetries among investors, improving the liquidity of firms' shares (e.g., Glosten and Milgrom, 1985; Diamond and Verrecchia, 1991).² We document that our standardized measure is negatively associated with firms' bid-ask spreads. This association is robust to controlling for liquidity determinants (including firm size), accounting for secular trends (e.g., increasing internet usage),

¹For our 8-K filings measure we use the word count of voluntary 8-K filings. He and Plumlee (2020) suggest that this measure is superior to measures based on the count of filings.

 $^{^{2}}$ See Beyer et al. (2010) for a thorough review of this literature.

and focusing on within-firm changes. Notably, it even holds when separately controlling for the established measures of voluntary disclosure. In sum, the liquidity result suggests that our standardized measure captures public information that is relevant for investors and incremental to the established measures of voluntary disclosure.

Lastly, we discuss three possible use cases of website-based disclosure measures to illustrate the broad applicability and promise of our measure for future research. We first document that our standardized measure can be used to study disclosures of small (and private) firms. While plentiful and of interest in their own right, these firms are often excluded from disclosure studies due to biased database coverage (e.g., Chuk et al., 2013). Our measure allows reducing the bias toward larger and publicly listed firms in disclosure research. We next document that our standardized measure can be used to create a comparable disclosure measure for international firms. As such, our website-based measure complements the annual-report-based measure of Lang and Stice-Lawrence (2015), facilitating international accounting research. Lastly, we document that our measure can be customized to study firms' content-specific disclosure changes (e.g., on ESG issues) around salient events.

Overall, our empirical evidence suggests that website-based disclosure measures provide useful proxies for firms' voluntary disclosure. They are particularly useful for studying disclosures of small, private, or international firms for which existing capital-market-specific disclosure measures are typically unavailable. Moreover, firms website disclosures allow creating customized, audience-specific disclosure measures and studying communication with audiences other than investors in public capital markets. This point is particularly important given that firms consider their various stakeholders and audiences when making public disclosure decisions (Crawford and Sobel, 1982; Newman and Sansing, 1993; Gigler, 1994; Goltsman and Pavlov, 2011); and given the recent push toward ESG disclosures and stakeholder capitalism (Grewal and Serafeim, 2020).

Our paper contributes to the literature on corporate disclosure. Despite the prominence of corporate websites as a communication channel for firms and information source for stakeholders, the literature provides primarily small-sample evidence on firms' disclosures via corporate websites. Recent examples include Pownall et al. (2015), who examine the link between the content of IR pages and liquidity for stocks traded on the Euronext, Arora and Lodhia (2017), who examine the role of website-based environmental disclosures at BP for reputation risk management purposes, and Saxton et al. (2014), who examine the association between website disclosures and charitable contributions at non-profit organizations.³ We contribute to this literature by providing a large-sample investigation of firms' disclosure via corporate websites and developing a novel website-based measure of voluntary disclosure.

Our paper further adds to the literature on information dissemination. Prior literature investigates the role of various information channels, including the media (Bushee et al., 2010; Rogers et al., 2016), EDGAR (Christensen et al., 2017; Rogers et al., 2017), Wikipedia (Zhu, 2019), and social media (Blankespoor et al., 2014; Blankespoor, 2018). We complement this literature by investigating one of the arguably most important channels through which firms communicate with their stakeholders: their corporate websites.

Lastly, our paper enriches the literature on the measurement of firms' voluntary disclosure. Established measures capture several dimensions of firms' voluntary disclosure, including the quantity (Dyer et al., 2017), disaggregation (Chen et al., 2015), frequency (Marquardt and Wiedman, 1998; He and Plumlee, 2020), and precision (King et al., 1990; Hutton et al., 2003) of disclosures such as SEC filings and management forecasts. Our website-based disclosure measure captures similar dimensions, but also complements the established measures in that it is more broadly available and targeted toward a broader audience than just capital-market participants.

Our paper is related to the growing literature making use of the Wayback Machine in various contexts. Several papers, for example, use the Wayback machine to obtain historic information on specific aspects, such as lists and locations of regulators' offices (e.g., Ege et al., 2020; Lim et al., 2016; Mas, 2017). Most closely related to our paper, a concurrent

³For other early examples on the role of firms' websites as a communication tool, see, for example, Koreto (1997), Ettredge et al. (2001), Matherly and Burton (2005), Cormier et al. (2009), and Pollach (2011).

working paper by Lynch and Taylor (2021) documents that publicly listed firms in the U.S., since 2008, can substitute their mandatory financial information disclosure on EDGAR with corresponding disclosures on their corporate websites. Consistent with our paper, they document that corporate websites contain useful information for investors. Distinct from Lynch and Taylor (2021), our paper is not limited to mandatory financial information and the capital-market context. Our exploration of corporate websites suggests that firms use their websites more broadly to communicate with a wide range of stakeholders. As such, we propose a novel and versatile measure of firms' voluntary public disclosure based on their websites' size and content.

2 Construction of Website Measure

2.1 Wayback Machine

We obtain full historical records of corporate websites from the Wayback Machine, a digital archive of websites which is maintained by the Internet Archive.⁴ Since 1996, the Internet Archive, a nonprofit organization financed by various foundations and individual donations, aims at preserving digital contents of all sort in a context of fast technological change. To this end, it has created a digital record of websites and other digital contents (e.g. softwares, movies, music...), similar to a regular print library. It currently maintains more than 45 Petabytes of data stored on various servers across the U.S. which, as part of its mission, the Internet Archive grants open access to.

As one of the main services of the Internet Archive, the Wayback Machine allows tracking the evolution of websites over time. Its algorithm periodically crawls the internet to keep a record of all existing and newly created websites. The resulting records are snapshots of almost all existing websites at different points in time, accessible via the Wayback Machine's user-friendly interface. This interface prompts users to enter a web URL and allows them

⁴https://archive.org/web/

to navigate into the respective website as it stands at different points in time (see the time banner on the screenshots in Appendix A).

In addition to the user-friendly interface, the Wayback Machine provides access to the underlying archive data through an Application Programming Interface (API) which allows collecting the historical structure of the website. For each archived website, the API provides access to the list of all URLs constituting the website (variable 'urlkey'), a 'timestamp' variable indicating the day at which an URL exists, the size (in octets or bytes) of each URL, and a 'status code' variable indicating whether an URL points to a webpage or if the query results in an error (e.g. 404 for the '404 not found' error). In addition, the 'mimetype' variable indicates whether a specific URL contains text only (e.g. 'text/html'), or enriched content such as images, videos, or specific applications (e.g. Excel files, PDF files).

2.2 Sample

To construct our sample, we start with all firms in COMPUSTAT U.S. with non-missing corporate website URLs (COMPUSTAT variable 'weburl'). This step results in 13,521 firms. We collect historical data on each website (an archive) from the Wayback Machine API by using queries of the following form: http://web.archive.org/cdx/search/cdx?url=www.cecoenviro.com&matchtype=domain&collapse=timestamp:10&matchType=prefix&output=json

In the above example query, we ask the Wayback Machine API for all records of the website of CECO Environment. We obtain historical records from cecoenviro.com (the main domain) as well as all its subdomains (by setting 'matchtype=domain'). This query returns a list of records (rows), each one representing an URL. The results are provided in JSON format, a light format that is well suited for the storage of large databases. Following these steps, we are able to download 13,260 archives of corporate websites (98% of the initial dataset), amounting to about 155 Gigabytes of data.

We read each JSON file and remove URLs whose 'status code' corresponds to an error and

we extract the date of each record from the 'timestamp' variable. A given URL may appear several times in the database, corresponding to each time the website has been crawled by the Wayback Machine algorithm. We first collect all the records available for each firm. As we are interested in creating measures of voluntary disclosure on a quarterly basis, we compute the average values for that URL in a given quarter whenever there are several records of an URL available in a given quarter.

Finally, we keep firms for which we can compute capital-market variables (i.e., quarterly bid-ask spreads, market values, share turnover, and return variability) by merging our dataset with the CRSP-COMPUSTAT merged database. This step results in a sample of 5,591 unique firms corresponding to 209,186 firm-quarter observations over the 1997-2018 period. We add four commonly used measures of disclosures to our sample: the number of management earnings forecasts (Management Forecast, period: 2000-2018), the number of disaggregated items in firms' financial statements (Disclosure Quality, period: 1979-2017), the word count of voluntary 8-Ks filings (Voluntary 8-K filings, period: 2005-2016), and the number of characters in the original 10-K filings (gross 10-K file size, period: 1993-2017).

2.3 Measures

2.3.1 Website Size

We construct our main measure of website-based disclosure by taking the logarithm of firms' quarterly website size (*Website Size*: the total size of the website in bytes). This simple measure allows for a standardized comparison across firms and over time. Conceptually, it captures the quantity of information or level of detail provided via firms' websites, in line with previous voluntary disclosure measures (Loughran and McDonald, 2014; Chen et al., 2015; He and Plumlee, 2020). While firms can elect to provide mandatory disclosures (e.g., SEC filings) on their websites (e.g., Lynch and Taylor, 2021), the vast amount of variation in firms' website size can be expected to represent firms' voluntary disclosure choices.

To illustrate the variation captured by our website-based disclosure measure, we present

two examples of corporate websites in Appendix A. We show screenshots of the landing pages of two firms' websites as captured in Q4 of 2017. The two example firms, *Celcuity Inc.* and *Teligent Inc.*, operate in the same industry (i.e., pharmaceuticals), are listed on the NASDAQ and belong to the same (third) decile by market capitalization, to facilitate comparison. According to our standardized measure, *Celcuity Inc.* had a relatively small website in 2017 with a size of 449 Kilobytes. The website is fairly generic and contains only a brief description of firms' products and standard financial information. By contrast, the *Teligent Inc.* website, according to our standardized measure, was much larger in 2017, with a size of 11,294 Kilobytes. It contained detailed information on various topics, including the firm's investment, manufacturing capabilities, and existing products. The "investing" section further contained a document detailing their plan for expansion and the job positions associated with this expansion. Browsing the two examples of historical websites illustrates that the relative size of the websites tends to reflect the amount of information provided via these sites.

2.3.2 Website Content

We complement our standardized website-size measure with measures examining the content of websites. The content-based measures aid our understanding of the information disclosed on firms' websites and illustrate the potential for customization of our website-based measure for specific research questions (e.g., about product information or ESG issues). To construct content-based measures, we take three distinct approaches.

First, we construct disaggregated measures capturing the components (e.g., text, applications, image, video) of each website and the used file formats (e.g., PDF, Word, Excel, Java) using the 'mimetype' information. This disaggregation of firms' websites into components and file formats allows examining which specific parts of firms' website provide outsiders (e.g., investors) with useful information.

Second, we construct broad measures of website contents by parsing the words embedded

in the URLs of firms' websites. Each URL in our database is a string that is potentially informative about the content of the webpage to which it points. For instance, an URL that contains the keyword "investor" is likely to point to the investor relation section of the website, while an URL containing the word "careers" points to job opportunities for prospective job applicants. We decompose the URLs into relevant content- and/or audiencerelated categories as follows. We first purge all URL strings of the website host name. We next retain all words in the URL appearing in a standard English dictionary.⁵ After additionally purging stop words (e.g. pronouns), we obtain a list of 92,968 unique words. For each word, we then compute the frequency with which firms are using it in their URLs. This distribution of word frequencies is highly skewed, as several words are only used once (e.g., there are 23,020 words that are only used once). The words with low frequencies are typically highly specific words that are idiosyncratic to a firm's business (e.g. cyclotron, photopolymer, or zebrawood). The most frequent words, by contrast, tend to be programming-related terms that convey very little content-specific information (e.g. txt, robots, or contact). We remove words in the top 10% (used by more than 4.972 firms) and bottom 10% (used by less than 562 firms) of the distribution to focus our manual classification of URL content into relevant categories, leaving us with 1,701 unique words to classify.⁶ We classify the words as belonging to broad content categories (e.g., "Investor and Corporate Governance") identified in two early exploratory studies on disclosures on corporate websites (Aerts et al., 2006; Cormier et al., 2009).^{7,8}

Third, we construct specific measures of website contents by analyzing the text on firms'

⁵We use the following dictionary: https://github.com/dwyl/english-words

⁶We do not rely on statistical methods (e.g. Latent Semantic Analysis) to group words into categories as we are not interested in words that co-occur, but rather seek to understand what broader content or audience a given word may belong to or speak to.

⁷Aerts et al. (2006) categorize the disclosed content of websites of 56 European and North American firms. Cormier et al. (2009) adopt the same scheme to categorize website content of 189 listed Canadian firms.

⁸We are able to manually classify 502 words out of the full list of unique words into the broad content (or audience) categories. The remaining 1,199 words cannot be used to identify the topic of the URL either because they are programming-specific (e.g. background, banner, color), or because they do not convey any information alone (article, chart, deliver, etc...).

websites. For these specific measures, we need to query the Wayback Machine API to obtain all records of websites corresponding to the firms' URLs, not just the meta data. Using the 'timestamp' variable, we keep the last-seen record of each firm's URL for each calendar quarter. This gives us a snapshot of corporate websites as they stand at the end of each calendar quarter. To analyze the content of these websites, we then proceed as follows: We first parse the host URL (i.e. the website main page) as well as its sub-URLs, and store their content in a vector.⁹ We next clean the resulting parsed characters: we remove irrelevant textual contents (e.g. HTML tags, Java scripts), numbers, and punctuation marks. We finally stem each word using the Porter algorithm.¹⁰ The resulting dataset comes in the form of a list of words constituting each website in a given quarter. Equipped with those data, we use a bag-of-words approach to examine the frequency of specific words/content of interest.

3 Variation of Website Measure

Following our website-based measure construction, we describe its distribution, explore the content of the underlying websites, examine the coverage of our measure, and decompose its variance in this section.

3.1 Descriptive Statistics

Table 1 Panel A provides descriptive statistics for our standardized disclosure measure (*Website Size*), the four established disclosure measures, and capital-market measures. The table documents that the average website size of our sample firms is about 12 Megabytes. Its distribution is highly right skewed, as is typical for size variables. To adjust for the skewness and reduce the impact of outliers, we use the natural logarithm of website size as

⁹For efficiency, we design our web-parsing program to stop either after parsing the third layer of sub-URLs or after parsing at least 100 sub-URLs.

¹⁰The Porter stemming algorithm removes the common morphological and inflectional endings from words in English. For example, the Porter algorithm reduces *argue*, *argued*, *argues*, and *arguing* to the stem *argu*.

our standardized measure of voluntary disclosure. Figure 1 plots this (logged) distribution. It shows that our measure exhibits substantial variation and that its distribution closely follows a bell shape.

3.2 Content

Figure 2 plots the components and file formats making up our sample firms' websites, and their evolution over time. The upper panel documents that, for most of our sample period, firms' websites are primarily constituted of text. Over time, the importance of images has increased though. As for file formats, the lower panel documents that, for most of our sample period, PDF and/or Java files are the dominant file formats used on firms' websites.

Figure 3 plots the relative frequency of four distinct content categories by industry (defined using GICS codes) for a cross-section of our sample firms' websites as of the last quarter of 2016. The categories are assigned based on the URL string content. The "Product, Strategy and Processes" category comprises URLs with content on firms' production, products, innovation, customers, and suppliers. The "Geography" category comprises URLs with information on firms' locations (domestic and foreign), which is useful for understanding the scope of firms' activities. The "Investor and Corporate Governance" category comprises URLs with information on firms' financials and corporate governance. Finally, the "Human Resources" category encompasses URLs with content related to human capital, including job opportunities.¹¹

Three broad patterns emerge from the figure. First, the "Investor" and "Product" categories largely dominate in almost all industries. Second, we observe significant variation in the relative share of each of the four categories across industries. Third, the relative share of the four topics can be linked to certain industry characteristics, such as the nature of their

¹¹In addition, we categorize words in further categories: i) "Communication", which encompasses all items related to the press, social media, events, public relations, and announcements; and ii) "Corporate Social Responsibility", which encompasses words related to environment, diversity, and social impact. As those two categories represent only a small portion of firms' websites, we exclude them from the figure for visualization purposes. The full list of words used for each topic is shown in Appendix B.

assets, the geographical scale of their operations, or the importance of labor.

Industries with a high share of the "Product, Strategy and Processes" category, for example, tend to be industries with complex assets or products (e.g. energy, utilities, telecommunication services), industries in which firms communicate on their innovations, and consumerfacing industries. For instance in the Utilities industry, Alliant energy (Nasdaq: LNT) dedicates a large fraction of its website to its innovative energy solutions. Similarly, in the Telecommunication Service industry, Globalstar (Amex: GSAT) describes its telecommunications products revolving around its satellite network at length. By contrast, industries with a low share of the "Product" category tend to be industries with well-known products. General Mills (NYSE: GIS) provides a stark example. On its website, it simply lists its brands, most of them being household names in the U.S., on one sub-URL. As those brands are advertised via different channels (e.g., via TV ads or retailers' in-store promotions), the website of General Mill mostly acts as a medium to communicate financial information about the conglomerate to investors. Consistent with this idea, Figure 3 shows that firms in the "Consumer Durables", "Consumer Services", or "Food, Beverage, and Tobacco" industries have a low share of the "Product" category, but a relatively high share of the "Investor" category.

Compared to the "Product" and "Investor" categories, the "Geography" and "Human Resources" categories are less frequent. Still, there is notable variation in the share of these categories across industries, and the information contained in those categories promises to be of interest for future research (e.g., on location choice and labor demand). The "Geography" category, for example, is more pronounced in industries in which firms' location is a key strategic choice (e.g. Real Estate), or in which local distribution matters (e.g., Consumer Durable, Insurance). It can be used to gain insights on firms' decision on where to produce and sell their goods and services. The "Human Resources" category is more pronounced in the hospitality sector ("Consumer Services"), in industries that value skilled labor ("Software and Services", "Healthcare Equipment"), and in industries with mostly intangible assets ("Media"). It is widely driven by recruiting-related information (e.g., on careers and skill requirements) and tends to reflect firms' need for labor.

In sum, our content exploration suggests that corporate websites contain information on central aspects of firms' business. This information is relevant for various stakeholders, including investors, but also customers, suppliers, and employees. The extent to which certain aspects of firms' business are discussed appears to vary with salient industry characteristics. Importantly, while we document substantial across-industry variation in firms' website content, we stress that there is also significant cross-sectional variation between firms in a given industry. As such, the website content classification can be used in future research to examine the relative importance of various stakeholders across firms. For instance, the information in the "Human Resources" category could be used to measure firm-level investment in human capital (which can often not be gleaned from firms' financial statements), complementing existing measures such as those based on employee satisfaction scores (e.g., Edmans, 2011).

3.3 Coverage

Table 1 Panel B compares the coverage of our website-based measure with the coverage of established disclosure measures (management forecasts, voluntary 8-K filings, disclosure quality, and 10-K file size) over time. The comparison starts in 2005, which corresponds to the first year in which all five measures are available. Relative to all firm-years in the CRSP-COMPUSTAT universe, our measure's coverage is 54% in 2005 and reaches a high of 94% in 2018, at the end of our sample period. This gradual increase reflects the expanding coverage of the Wayback Machine over time. Moreover, it reflects that some firms' website domains (main URLs) may have changed over time. As COMPUSTAT only reports the latest website domain, we construct our measure using the latest domain as reported in 2019 which inevitably contributes to missing observations and limited coverage in the earlier sample period.¹² Notably, this issue can be circumvented going forward by using annual snapshots of firms' URLs provided on COMPUSTAT to track firms' future website changes.

The coverage of our measure, especially in the latter years, is remarkably high. It compares favorably to the coverage of the voluntary disclosure measures based on management forecasts and disclosure quality. The coverage of the 8-K based disclosure measure and the gross 10-K filing size are even higher than the coverage of our website-based measures though. This higher coverage, however, is in part driven by the fact that missing 8-K disclosures are coded as zeros. Moreover, it is due to the fact that both measures exploit variation across mandatory filings required by all the CRSP-COMPUSTAT firms in our sample. Outside of this sample (e.g., among private firms or firms listed in foreign capital markets), the coverage in contrast will be zero, whereas our website-based measures can be expected to cover a substantial fraction of firms outside of the CRSP-COMPUSTAT universe. See Section 5.2 for an exploration of our measure's coverage in a sample of international listed firms.

Table 1 Panel C documents that the size distribution of firms with non-missing websitebased measures closely resembles the distribution of the CRSP-COMPUSTAT universe over the 2005-2016 time period (i.e., the period with overlap for all disclosure measures). The average firm with non-missing website-based measures is only slightly larger than the average CRSP-COMPUSTAT firm, suggesting that the missing values primarily occur for the smaller firms in the earlier sample period. Unlike our website-based measure, the management forecast and disclosure quality measures are only available for a select group of firms. The management-forecast measure is primarily available for larger firms, whereas the disclosurequality measure is primarily available for smaller firms; likely because the latter does not cover financial institutions.¹³ Unsurprisingly, the 8-K and 10-K measures, which are built off of mandatory filings, follow a size distribution similar to that of our CRPS-COMPUSTAT

 $^{^{12}\}mathrm{We}$ suspect that mergers and acquisitions also contribute to the low coverage at the beginning of our sample.

¹³Unlike the management-forecast measure, our measure is a continuous measure. Accordingly, it also tends to be available for smaller firms with limited disclosure incentives. Instead of taking the value of zero for these firms, our measure will show a lower website size as long as they have a website.

universe.

Taken together, Panels B and C of Table 1 document that our website-based measure is widely available and exhibits substantial coverage advantages over several of the established measures of disclosure.

3.4 Variance Decomposition

Table 1 Panel D explores sources of variation making up our website-based measure. In particular, the panel dissects systematic factors driving variation in our disclosure measures. It documents that common industry and time factors only explain a relatively small amount of our measures (about 4% and 12%, respectively). The latter is particularly noteworthy. It suggests the variation in our website-size measure does not merely reflect a time trend toward greater internet and website use. In contrast to common industry and time factors, fixed firm factors explain a large part of our disclosure measures (about 42%). Regressions of our disclosure measure on firm fixed effects explain roughly four (ten) times more variation in our disclosure measures than regressions using time (industry) fixed effects. This feature supports the notion that our measure captures systematic variation in firms' voluntary disclosure decisions. These decisions tend to be sticky due to a commitment to a particular disclosure strategy (Leuz and Verrecchia, 2000). Notably though, there is still substantial variation left unexplained even after accounting for fixed firm and time factors (i.e., 100% - 57% =43%). Accordingly, our measure promises to also lend itself to event study or difference-indifferences research designs exploring differential changes in firms' disclosures over time. See Section 5.3 for an event study of firms' disclosure around a major news event.

4 Validation of Website Measure

To formally validate our website-size measure as a measure of voluntary disclosure, we examine its association with proxies for two important constructs: firms' voluntary disclosure incentives and investors' information asymmetry. As a proper measure of voluntary disclosure, our website-size measure should be positively related to established proxies capturing firms' voluntary disclosure decisions. Accordingly, we examine the association of our measure with voluntary disclosure proxies established in the literature, including management forecasts, voluntary 8-K filings, disclosure quality, and gross 10-K file size (e.g., King et al., 1990; Loughran and McDonald, 2014; Chen et al., 2015; He and Plumlee, 2020). As a proper measure of voluntary disclosure, our website-size measure should further be negatively related to proxies for investors' information asymmetry and firms' share illiquidity. Accordingly, we examine the association of our measure with bid-ask spreads, the main proxy for investors' information asymmetry and firms' share liquidity established in the literature (e.g., Lesmond et al., 1999; Leuz and Verrecchia, 2000; Brüggemann et al., 2018).¹⁴

4.1 Voluntary Disclosure

Table 1 Panel E presents univariate correlations between our website-size measure and the four established voluntary disclosure measures. Our website-size measure is positively correlated with all four established disclosure measures, providing *prima facie* evidence in support of the validity of our empirical proxies as measures of firms' voluntary public disclosure. To strengthen this evidence, we next turn to multivariate regressions, allowing us to account for different sets of fixed effects (e.g., time, industry, or firm effects).

Table 2 presents results from regressions of our website-size measure on the four established voluntary disclosure measures. In our baseline specifications with year fixed effects, we find that all four voluntary disclosure measures are significantly positively associated with our website-size measure (columns 1, 4, 7, 10). When additionally accounting for industry fixed effects, these associations attenuate for the relation with the management-forecasts measure and the disclosure-quality measure (columns 2 and 5), whereas they remain stable

 $^{^{14}}$ For brevity, we only tabulate results using bid-ask spreads. In untabulated results, we find that our results are widely robust to using a liquidity factor as in Christensen et al. (2013) that aggregates the following three liquidity proxies: zero returns, the Amihud illiquidity measure, and bid-ask spreads.

for the relation with voluntary 8-K filings and gross 10-K file size (columns 8 and 11). When accounting for fine firm instead of broad industry fixed effects, the association between our website-size measure and the established disclosure measures remains positive and significant for all measures but the disclosure-quality measure (column 6). The insignificant relation between our website-size measure and the disclosure-quality measure upon inclusion of firm and year fixed effects likely reflects the limited time-series variation in the disclosure-quality measure. By contrast, the significant relationship between our website-size measure and management forecasts, voluntary 8-K filings, and gross 10-K file size, even when only focusing on differential changes of firms' disclosures over time, plausibly indicates that our website-size measure captures changes in firms' voluntary disclosure decisions.

The stable associations suggest our website-size measure captures similar variation as established disclosure measures, validating our measure as a useful proxy for firms' voluntary disclosure. The limited power of the established disclosure measures (and fixed effects) to explain our measure's variation (e.g., R^2 s range from 7% in column 7 to 57% in column 9), however, also suggests our website-based measure captures additional, independent variation. This variation may be related to disclosures geared toward stakeholders other than capitalmarket investors, the main audience of the four existing voluntary disclosure measures. This feature allows studying distinct audiences (e.g., customers or regulators) and topics (e.g., ESG instead of financial performance), adding to the incremental usefulness of our measure over and above existing measures. See Section 5.3 for an exemplary application of our website-based measure to an ESG related disclosure event.

4.2 Liquidity

Table 3 presents results from regressions of bid-ask spreads on our website-size measure. The unconditional association (no controls or fixed effects) of our website measure with firms' bid-ask spreads is significantly negative (column 1; see also Table 1 Panel E). This association attenuates, but remains negative and significant after controlling for liquidity determinants drawn from (Christensen et al., 2013), including firm size (i.e., market value) (column 2). This pattern suggests that our measure, while positively correlated with firm size, does not merely reflect firm size. The association further attenuates, but remains negative and statistically significant after controlling for time effects (column 3), industry effects (column 4), and firm effects (column 5). These results suggest that our website-size measure is robustly negatively associated with firms' bid-ask spreads, as is expected of a valid measure of voluntary disclosure.¹⁵ Notably, this relationship holds even when accounting for time trends (e.g., increased internet usage), controlling for cross-industry differences (e.g., business-to-business vs. business-to-consumer industries), and focusing on firm-level changes over time (e.g., added website content). It also holds when defining our website-based measure more narrowly as the text-only website size (i.e., after purging pictures and other applications), as reported in Table 4.

Table 5 presents additional results from regressions of bid-ask spreads on our website-size measure *conditional* on each of the established disclosure measures. The conditional associations allow us to examine whether our website-size measure is subsumed by any of the established measures or provides incremental information relevant for investors.¹⁶ In Panel A, we document that both our website-size measure and the management-forecast-based measure are negatively associated with bid-ask spreads. After controlling for liquidity determinants and various sets of fixed effects, the associations of both our website-size measure and management forecasts with bid-ask spreads drop in size but remain statistically significant. In Panel B, our website-size measure remains negatively correlated with bid-ask spreads

¹⁵The negative association is also evident in the sub-sample period before the SEC allowed firms to disclose financials via firms' websites instead of via EDGAR (except in the most demanding firm and time effects specification). This robustness suggests the negative relation between our website-size measure and firms' liquidity does not merely reflect firms' disclosure of mandatory filings via their websites (Lynch and Taylor, 2021).

¹⁶We set missing values of the management-forecast and disclosure-quality measures to zero (within the sample period during which each of these measures is generally available). This approach allows us to examine the incremental impact of our website-size measure, while limiting the impact of sample-size restrictions (associated with the distinct measures) on the website-size coefficient estimate. The number of observations nevertheless varies across the four panels as a result of the distinct frequencies (e.g. disclosure-quality is an annual measure) and time periods (8-K for instance is available only after 2005) for which the four established measures are available. See Table 1 for details on the availability of each measure.

throughout all five specifications when including the disclosure quality variable. In contrast, the disclosure quality measure displays an unstable relationship with bid-ask spreads, with the sign of the association flipping depending on the set of fixed effects. In Panel C, we add the voluntary 8-K filings measure from He and Plumlee (2020). Our website-size measure is negatively and significantly associated with bid-ask spreads in all specifications except the specification with industry and time fixed effects. The association between the voluntary 8-K filings measure and bid-ask spreads remains systematically negative and statistically significant at conventional levels. Finally, in Panel D, we include the gross 10-K filings measure developed by Loughran and McDonald (2014). Our website-size measure remains negatively and statistically significantly associated with bid-ask spreads across all five specifications. The gross 10-K filings measure, by contrast, flips signs depending on the sets of fixed effects.

In sum, our website-size measure's stable associations with firms' bid-ask spreads suggest that our measure is negatively related to firms' share illiquidity, consistent with theory (Verrecchia, 2001) and prior voluntary disclosure measures (Coller and Yohn, 1997; Chen et al., 2015; He and Plumlee, 2020). The negative relation appears to emerge because our website-based measure captures both similar dimensions as established voluntary disclosure measures, but also incremental information relevant for investors. As such, our our measure can be used varied settings. It can be used in capital-market contexts in combination with established measures. In such contexts, it can also be used instead of established measures; for example, if those measures are unavailable (e.g., for small or international firms; see sections 5.1 and 5.2). In addition, our measure promises to also capture information relevant for other audience (see section 5.3 for an example of environmental disclosure). Accordingly, its applicability is not limited to capital-market contexts. We discuss exemplary applications of our website-based disclosure measure in the next section.

5 Applications of Website Measure

Following the validation of our website-based disclosure measure, we discuss three potential applications of our standardized and customized website-based measures to showcase their broad applicability.

5.1 Small and Private Firms

Our website-based disclosure measure promises to be useful in studying disclosures of small firms and private firms. For these firms, established disclosure measures are often missing. The missing disclosure measures for small and private firms in part reflect the fact that these firms tend to exhibit lower public disclosure incentives (e.g., Buzby, 1975; Breuer et al., 2020) The missing measures, however, also reflect the fact that for small and especially private firms, there is usually no centralized disclosure platform (e.g., EDGAR) or database which allows researchers to easily measure firms' disclosure activity (e.g., Chuk et al., 2013).

Our website-based measure, by contrast, can be constructed for small and private firms, as many databases (e.g., Orbis, Dun & Bradstreet), while often lacking financial information, at least contain small and private firms' website URL. Absent centralized disclosure platforms and dedicated information intermediaries (e.g., analysts, credit bureaus), small and private firms can be expected to rely on their corporate websites to communicate with the public. In support of this argument, Bourveau et al. (2021), building on our website-based disclosure measure, document that U.S. private firms inform interested stakeholders (e.g., potential investors and employees) about their business growth via industry awards (e.g., INC 500 awards) published on their websites. This finding stands in contrast to the widely-held belief that U.S. private firms do not provide any financial information publicly (e.g., Minnis, 2011; Lisowsky and Minnis, 2020).

Our measure not only allows studying private firms' disclosures, but can also help reduce bias towards larger firms in disclosure research. To showcase this benefit, we re-examine the association between our website-size measure and firms' liquidity (detailed in section 4.2) separately for listed firms of different sizes. We do so because, since the discontinuation of the AIMR ratings, the literature has primarily relied on management forecasts as a measure of voluntary disclosure (e.g., Beyer et al., 2010; Balakrishnan et al., 2014).¹⁷ As documented in Panel C of Table 1, the use of management forecasts tends to skew the sample of firms toward larger firms for reasons related to firms' disclosure incentives and data providers' coverage decisions (Chuk et al., 2013). By contrast, our measure allows examining voluntary disclosure not just for large listed firms, but also for small cap firms. This feature permits generating more generalizable findings applying to the entire spectrum of public firms.

Figure 4 plots the association of our website-based measures with bid-ask spreads for each firm-size (market-capitalization) decile. Three noteworthy patterns emerge. First, our website-size measure is systematically negatively associated with bid-ask spreads for all but the largest firms, consistent with website disclosures reducing information asymmetry for almost all firms. Second, while negative, the point estimate for the smallest firms (first decile of market capitalization) is small, presumably because this subgroup of firms is highly illiquid. Third, the negative relation is most pronounced among the remaining lower deciles of market capitalization (deciles 2 to 5).¹⁸ This pattern is consistent with these smaller firms exhibiting higher information asymmetry and poorer coverage by traditional information intermediaries (e.g., the media and analysts).¹⁹ As a result, disclosures via websites seem to play a prominent role for these firms in efforts to alleviate information asymmetries among capital-market participants. By contrast, the point estimates are smaller for the largest firms in U.S. capital markets, indicating that variation in firms' websites plays a more modest role in alleviating information asymmetry for firms with a rich information environment.

¹⁷The length of the MD&A section or the number of voluntary 8-K filings are so far primarily used to complement results obtained using management earnings forecasts.

¹⁸The estimate for the 9th decile presents an exception from this rule. It may be a chance result given its sensitivity to research design choices (e.g., it is absent when using (unlogged) bid-ask spreads).

¹⁹Fang and Peress (2009) document that firm size has an strong positive effect on media coverage of listed firms. Similarly, the literature has long established that small cap firms received much lower coverage from equity analysts (e.g., Rajan and Servaes, 1997).

5.2 International Firms

The vast majority of the disclosure literature focuses on U.S. listed firms. Accordingly, we validate our website-based disclosure measure for a sample of U.S. listed firms. Importantly though, our measure is not specific to these firms, unlike voluntary 8-K filings, for example. It can be constructed for any firm with a corporate website. As such, it, for example, promises to enable international and cross-country research on firms' disclosure.

International accounting research is a fast growing stream of the accounting literature. To date, it has amassed substantial evidence on the extent and implications of comparability of accounting standards and numbers (e.g., Alford et al., 1993; Amir et al., 1993; Land and Lang, 2002; Ball et al., 2003; Leuz et al., 2003; Barth et al., 2008; Daske et al., 2008; Barth et al., 2012; Ball, 2016). Recently, Lang and Stice-Lawrence (2015) extended the international accounting literature by applying textual analysis approaches to measure the quantity and quality of the information contained in international firms' annual reports.²⁰ Inspired by this recent development, our website-based measure provides an additional measure of disclosure quantity and content for international firms.

To highlight the promise of our website-based measure for international research, we examine the coverage of international firms' websites in the Wayback Machine. Using COM-PUSTAT Global, we obtain information on international (non-U.S.) firms, including their website URL. We match those data to the Wayback Machine data via the firms' URL. Table 6 lists the unique number of firms (in 2015) for each country covered in COMPUSTAT Global (with 10 or more unique firms) and the corresponding coverage (i.e., match percentage) in the Wayback Machine data. For the 93 countries in Table 6, the coverage is consistently high, most often over 97%. The high coverage documents that both our standardized website-size measure and customized website-based measures can be constructed for a comprehensive sample of international listed firms using the Wayback Machine's historical website informa-

 $^{^{20}}$ Exploiting the richness of annual reports for international firms, Stice-Lawrence (2017) examine the role of regulatory monitoring using measures derived from firms' annual reports.

tion.

5.3 Content-Specific Event Analysis

Our website-based disclosure measure can be customized to capture specific information relevant in a given research context (see section 3.2). To showcase the promise of customized website-based measures, we examine U.S. public oil & gas firms' disclosure response to the Deepwater Horizon (BP) oil spill in 2010, a salient news event.

The Deepwater Horizon oil spill constitutes the largest marine oil spill in history. It was caused by an explosion on the Deepwater Horizon oil rig, operated by British Petroleum (BP) and located in the Gulf of Mexico on April 20, 2010, approximately 41 miles off the coast of Louisiana. As a result of the explosion and the ensuing collapse of the oil rig's structure, 4.9 million barrels of oil were spilled into the ocean. This environmental disaster received considerable coverage in the media and prompted public and private activism, including law suits by the U.S. government and activism campaigns by NGOs.

The oil spill setting allows us to examine an event-specific change in US public oil & gas firms' disclosure of environmental issues and risks. Given the increased awareness about the environmental risk associated with offshore drilling after the oil spill, we expect that oil & gas firms expand their discussion of environmental issues and risks on their websites (Leuz and Schrand, 2009; Bonetti et al., 2018). While firms in other industries may also face an elevated demand for environmental risk disclosures, we expect the impact to be strongest for oil & gas firms. To examine the differential impact on oil & gas firms, we compare their disclosure response to the response of firms operating in the pharmaceutical industry. The firms in the pharmaceutical industry tend to grapple with environmental risks too. The public awareness of those risks, however, is less likely to have increased substantially through the oil spill, compared to the awareness of the risks of oil & gas firms.

To construct a customized website-based disclosure measure for the oil spill event, we query and parse the entire content of firms' websites (as described in section 3.2). We process the website content of 175 treatment firms, composed of U.S. public oil & gas firms (SIC 13) with non-missing corporate website information in COMPUSTAT, and 300 control firms, randomly selected from the pharmaceutical industry (SIC 49).²¹ Using a bag-of-words approach (Loughran and McDonald, 2016), we calculate the relative frequency of the word 'environment' to measure firms' environment-related disclosures.²²

Figure 5 plots the frequency of environment mentions on oil & gas firms' websites and pharmaceutical firms' websites over time. The blue line corresponds to the frequency of mentions by oil & gas firms. It exhibits an upward trend with a visible jump immediately following the BP oil spill, demarcated by the vertical dashed green line. By contrast, the frequency of environment mentions by pharmaceutical firms is generally lower and remains rather flat over time. The visual inspection of Figure 5 provides *prima facie* evidence consistent with oil & gas firms significantly increasing their environmental disclosures on their websites around the unexpected BP oil spill relative to the benchmark firms.

To strengthen our inferences, we explicitly examine whether the time series of environment mentions by oil & gas firms and pharmaceutical firms exhibit structural breaks. To estimate the structural breaks, we follow the method developed by Bai and Perron (1998) and the implementation guidelines detailed in Bai and Perron (2003). Our structural break test endogenously estimates break points in the time-series data, without imposing any prespecified event or break dates. For the pharmaceutical firms, our test does not yield evidence supporting any structural breaks. By contrast, for oil & gas firms, our test uncovers evidence of two structural breaks, represented by the vertical dashed dark lines. Notably, the second break occurs immediately after the BP oil spill, resulting in an overlap of the line demarcat-

²¹Querying and parsing firms' entire website content is more time-intensive than creating our standardized measure. Accordingly, we restrict our sample to a limited number of firms operating in an industry which, like the oil & gas industry, also faces environmental risks.

 $^{^{22}}$ We compute the relative frequencies of all stemmed words for each website-quarter observation. To control for noise arising from URLs not properly archived in the Wayback Machine (especially in the earlier part of our sample period), we apply the following conditions to each website-quarter observation: we exclude all words that occur only once; we set the term frequency to missing if the total number of words on the website is less than 2,000; we set the term frequency to missing if its value is zero in that firm-quarter but nonzero in the previous and subsequent quarters; we winsorized term frequencies for all words at the 2.5% and 97.5% level for each firm-quarter to reduce the influence of outliers.

ing the oil spill event and the dashed line demarcating a structural break. This structural break, unlike the first one, preceding the Deepwater Horizon catastrophe by several years, is tightly identified, as indicated by the narrow confidence intervals (shown as red lines on the x-axis). It support the visual impression that oil & gas firms significantly increased their environment-related disclosures on their websites in response to the Deepwater Horizon oil spill.

This exemplary event study around the Deepwater Horizon catastrophe showcases the usefulness of customized measures of disclosure based on *specific* contents discussed on firms' websites for studying wide-ranging questions. It specifically highlights that firms discuss various topics on their websites, including environmental, social, and governance (ESG) issues; and that firms update their websites in a timely manner. The former allows researchers to not only measure firms' financial disclosures, but also construct novel disclosure measures capturing ESG and other non-financial information. The latter suggests that our website-based measures of disclosure can not only be used to study broad sample correlations between standardized disclosure measures and firm outcomes, but also to examine specific disclosure changes around relevant corporate events.

6 Conclusion

Corporate websites represent an important channel through which firms communicate with the public. Based on this insight, we construct and validate a novel measure of firms' voluntary disclosure using firms' historic website sizes and contents, archived by the Wayback Machine.

Our website-based measure of voluntary disclosure exhibits rich cross-sectional and timeseries variation in firms' website size and content. The size and content differences across firms and changes over time are positively associated with established measures of firms' voluntary disclosure (e.g., management forecasts and voluntary 8-K disclosures) and liquidity. These associations suggest our measure captures similar investor-relevant information as established measures of voluntary disclosures. Our measure, however, is not subsumed by any of the established measures. By contrast, it provides incremental information relevant to investors and other audiences (e.g., customers and employees) and is more broadly available than most of the established measures.

Our measure promises to be of use in various research settings. Our standardized websitesize measure, for example, can be used in broad-sample studies on firms' public disclosure, while customized content-specific measures can be constructed to study specific disclosure responses around corporate events. Our website-based measures hold particular promise for studying disclosures of small, private, or international firms for which comparable disclosure measures are lacking; for examining disclosures to audiences other than investors (e.g., customers and employees); and for researching disclosures regarding non-financial information (e.g., ESG disclosures).

References

- Aerts, W., D. Cormier, I. Gordon, and M. Magnan, 2006, Performance disclosure on the web: An exploration of the impact of managers' perceptions of stakeholder concerns, *The International Journal of Digital Accounting Research* 6, 159–194.
- Alford, A., J. Jones, R. Leftwich, and M. Zmijewski, 1993, The relative information content of accounting disclosures in different countries, *Journal of Accounting Research* 31, 183– 223.
- Amir, E., T. Harris, and E. Venuti, 1993, A comparison of the value-relevance of U.S. versus non-U.S. GAAP accounting measures using form 20-F reconciliations, *Journal of Accounting Research* 31, 230–264.
- Arora, M., and S. Lodhia, 2017, The BP Gulf of Mexico oil spill: Exploring the link between social and environmental disclosures and reputation risk management, *Journal of Cleaner Production* 140, 1287–1297.
- Bai, J., and P. Perron, 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47–78.
- Bai, J., and P. Perron, 2003, Computation and analysis of multiple structural change models, Journal of Applied Econometrics 18, 1–22.
- Balakrishnan, K., M. Billings, B. Kelly, and A. Ljungqvist, 2014, Shaping liquidity: On the causal effects of voluntary disclosure, *Journal of Finance* 69, 2237–2278.
- Ball, R., 2016, Why we do international accounting research, *Journal of International Accounting Research* 15, 1–6.
- Ball, R., A. Robin, and J. S. Wu, 2003, Incentives versus standards: properties of accounting income in four East Asian countries, *Journal of Accounting and Economics* 36, 235–270.
- Barth, M., W. Landsman, and M. Lang, 2008, International accounting standards and accounting quality, *Journal of Accounting Research* 46, 467–498.
- Barth, M., W. Landsman, M. Lang, and C. Williams, 2012, Are IFRS-based and U.S.GAAPbased accounting amounts comparable?, *Journal of Accounting and Economics* 54, 68–93.
- Beyer, Anne, Daniel A. Cohen, Thomas Z. Lys, and Beverly R. Walther, 2010, The financial reporting environment: Review of the recent literature, *Journal of Accounting and Economics* 50, 296 – 343.

- Blankespoor, E., 2018, Firm communication and investor response: A framework and discussion integrating social media, *Accounting, Organizations and Society* 68-69, 80–87.
- Blankespoor, E., G. Miller, and H. White, 2014, The role of dissemination in market liquidity: Evidence from firms' use of Twitter, *The Accounting Review* 89, 79–112.
- Bonetti, P., C. Cho, and G. Michelon, 2018, Environmental disclosure and the cost of capital: Evidence from the Fukushima nuclear disaster, Working Paper (available at: SSRN link).
- Botosan, C., 1997, Disclosure level and the cost of equity capital, *The Accounting Review* 72, 323–349.
- Bourveau, T., M. Breuer, and M. Muhn, 2021, The certification role of company awards for private firms, Working Paper.
- Breuer, M., K. Hombach, and M. A. Müller, 2020, The economics of firms' public disclosure: Theory and evidence, Working Paper (available at: SSRN link).
- Brüggemann, U., A. Kaul, C. Leuz, and I. Werner, 2018, The twilight zone: OTC regulatory regimes and market quality, *Review of Financial Studies* 31, 898–942.
- Bushee, B., J. Core, W. Guay, and S. Hamm, 2010, The role of the business press as an information intermediary, *Journal of Accounting Research* 48, 1–19.
- Buzby, S. L., 1975, Company size, listed versus unlisted stocks, and the extent of financial disclosure, *Journal of Accounting Research* 13, 16–37.
- Chen, S., B. Miao, and T. Shevlin, 2015, A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports, *Journal of Accounting Research* 53, 1017–1054.
- Christensen, H., E. Floyd, and M. Maffett, 2017, The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records, *Journal of Accounting and Economics* 64, 284–304.
- Christensen, H., L. Hail, and C. Leuz, 2013, Mandatory IFRS Reporting and Changes in Enforcement, *Journal of Accounting and Economics* 56, 147–177.
- Chuk, E., D. Matsumoto, and G. Miller, 2013, Assessing methods of identifying management forecasts: CIG vs. researcher collected, *Journal of Accounting and Economics* 53, 23–42.
- Coller, M., and T. Yohn, 1997, Management forecasts and information asymmetry: An examination of bid-ask spreads, *Journal of Accounting Research* 35, 181–191.

- Cormier, D., M.-J. Ledoux, and M. Magnan, 2009, The use of Web sites as a disclosure platform for corporate performance, *International Journal of Accounting Systems* 10, 1–24.
- Crawford, V., and J. Sobel, 1982, Strategic information transmission, *Econometrica* 50, 1431–1451.
- Daske, H., L. Hail, C. Leuz, and R. Verdi, 2008, Mandatory IFRS reporting around the world: Early evidence on the economic consequences, *Journal of Accounting Research* 49, 1085–1142.
- Diamond, D., and R. Verrecchia, 1991, Disclosure, liquidity, and the cost of capital, *Journal* of Finance 46, 1325–1359.
- Dranove, D., and G. Z. Jin, 2010, Quality disclosure and certification: Theory and practice, Journal of Economic Literature 48, 935–963.
- Dyer, T., M. Lang, and L. Stice-Lawrence, 2017, The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation, *Journal of Accounting and Economics* 64, 221– 245.
- Edmans, A., 2011, Does the stock market fully value intangibles? Employee satisfaction and equity prices, *Journal of Financial Economics* 101, 621–640.
- Ege, M., J. Glenn, and J. Robinson, 2020, Unexpected SEC resource constraints and comment letter quality, *Contemporary Accounting Research*, 37, 33–67.
- Ettredge, M., V. Richardson, and S. Scholz, 2001, The presentation of financial information at corporate Web sites, *International Journal of Accounting Information Systems* 2, 149– 168.
- Fang, L., and J. Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 2023–2052.
- Gigler, F., 1994, Self-enforcing voluntary disclosures, *Journal of Accounting Research* 32, 224–240.
- Glosten, L., and P. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.
- Goltsman, M., and G. Pavlov, 2011, How to talk to multiple audiences, Games and Economic Behavior 72, 100–122.

- Grewal, J., and G. Serafeim, 2020, Research on corporate sustainability: Review and directions for future research, *Foundations and Trends in Accounting:* 14, 73–127.
- He, J., and M. Plumlee, 2020, Measuring disclosure using 8-K filings, *Review of Accounting Studies* 25, 903–962.
- Hutton, A., G. Miller, and D. Skinner, 2003, The role of supplementary statements with management earnings forecasts, *Journal of Accounting Research* 41, 867–890.
- IR-Society, 2012, Best practice guidelines corporate websites updated march 2012, https://irsociety.org.uk/files/Best_Practice_Guidelines_Corporate_ websites_2012_FINAL.pdf, Accessed: 2019-07-15.
- King, R., G. Pownall, and G. Waymire, 1990, Expectations adjustments via timely management forecasts: Review, synthesis, and suggestions for future research, *Journal of Accounting Literature* 9, 113–144.
- Koreto, R., 1997, When the bottom line is online, Journal of Accountancy 183, 63.
- Land, J., and M. Lang, 2002, Empirical evidence on the evolution of international earnings, *The Accounting Review* 77, 115–134.
- Lang, M., and R. Lundholm, 1993, Cross-sectional determinants of analyst ratings of corporate disclosures, *Journal of Accounting Research* 31, 246–271.
- Lang, M., and L. Stice-Lawrence, 2015, Textual analysis and international financial reporting: Large sample evidence, *Journal of Accounting and Economics* 60, 110–135.
- Lesmond, D., J. Ogden, and C. Trzcinka, 1999, A new estimate of transaction costs, *Review* of Financial Studies 12, 1113–1141.
- Leuz, C., D. Nanda, and P. Wysocki, 2003, Earnings management and investor protection: An international comparison, *Journal of Financial Economics* 69, 505–527.
- Leuz, C., and C. Schrand, 2009, Disclosure and the cost of capital: Evidence from firms' response to the Enron shock, Working Paper (available at: SSRN link).
- Leuz, C., and R. Verrecchia, 2000, The economic consequences of increased disclosure, Journal of Accounting Research 38, 91–124.
- Lim, I., J. Hagendorff, and S. Armitage, 2016, Does distance impede regulatory monitoring? Evidence from the banking industry, Working Paper.

- Lisowsky, Petro, and Michael Minnis, 2020, The silent majority: Private u.s. firms and financial reporting choices, *Journal of Accounting Research* 58, 547–588.
- Loughran, T., and B. McDonald, 2014, Measuring readability in financial disclosures, *Journal* of Finance 69, 1643–1671.
- Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.
- Lynch, B., and D. Taylor, 2021, The information content of corporate websites, Working Paper (SSRN link).
- Marquardt, C., and C. Wiedman, 1998, Voluntary disclosure, information asymmetry, and insider selling through secondary equity offerings, *Contemporary Accounting Research* 15, 505–537.
- Mas, A., 2017, Does transparency lead to pay compression?, *Journal of Political Economy* 125, 1683–1721.
- Matherly, M., and H. Burton, 2005, An analysis of corporate website disclosures, Management Accounting Quarterly 6, 26–33.
- Minnis, M., 2011, The value of financial statement verification in debt financing: Evidence from private U.S. firms, *Journal of Accounting Research* 49, 457–506.
- Newman, P., and R. Sansing, 1993, Disclosure policies with multiple users, *Journal of Accounting Research* 31, 92–112.
- Pollach, I., 2011, The readership of corporate websites, *Journal of Business Communication* 48, 27–53.
- Pownall, G., M. Vulcheva, and X. Wang, 2015, The creation and segmentation of the Euronext stock exchange: A solution to the inadequacy of national securities regulators?, Accounting Horizons 29, 853–885.
- Rajan, R., and H. Servaes, 1997, Analyst following of initial public offerings, Journal of Finance 52, 507–529.
- Rogers, J., D. Skinner, and S. Zechman, 2016, The role of the media in disseminating insidertrading news, *Review of Accounting Studies* 21, 711–739.
- Rogers, J., D. Skinner, and S. Zechman, 2017, Run EDGAR run: SEC dissemination in a high-frequency world, *Journal of Accounting Research* 55, 459–505.

- Saxton, G., D. Neely, and C. Guo, 2014, Web disclosure and the market for charitable contributions, *Journal of Accounting and Public Policy* 33, 127–144.
- Stice-Lawrence, L., 2017, Too small to regulate: International evidence on regulatory monitoring, Working Paper (available at: link).
- Verrecchia, R., 2001, Essays on disclosure, Journal of Accounting and Economics 32, 97–180.
- Zhu, C., 2019, Investor demand for contextual information: Evidence from Wikipedia Working Paper.

Figure 1: Distribution of Website Size

The figure displays the distribution of our website-based measure of disclosure. *Website Size* is the size (in bytes) of the corporate website at the end of the quarter. There are 209,186 observations from 5,591 unique firms over the period 1997-2018. We trim *Website Size* at the 1st and 99t percentile before taking the natural logarithm.



Figure 2: Components of Websites Over Time

The figure charts the evolution of the four main components of corporate websites (top panel) and the evolution of applications used (bottom panel) over time. We compute firm-level fractions of each component at year-end as a proportion of the total website size (top panel) or as proportion of total application size (bottom panel). Firm-level fractions are averaged across firms year by year.





Figure 3: Content Categories

The figure presents the relative share of content categories of firms' websites based on their URL strings. We group firms into 25 industries based on their GICS code. For each corporate website in the last quarter of 2016, we purge each URL string from the website host name, stop words, and words that appear with either high frequency (i.e. in more than 90% of the websites) or low frequency (i.e. in less than 10% of the websites). We parse each resulting strings and classify words as belonging to one of the four content categories: *Product, Strategy and Processes, Geography, Investor and Corporate Governance*, and *Human Resources*. The words used for each category are listed in Appendix B. This classification allows us to assign URLs to specific content categories.



Figure 4: Website Size and Liquidity Across Firm Size

The figure plots the association between our website-size measure and firms' bid-ask spreads for each decile of our sample firms' size distribution (in terms of market capitalization). For each decile of market capitalization, we regress the natural logarithm of bid-ask spreads on the natural logarithm of *Website Size* and year-quarter dummies. The gray dots represent coefficient estimates and the gray whiskers represent the 95% confidence interval, calculated based on standard errors clustered at the firm level.



Figure 5: Environmental Disclosure Around the BP Oil Spill

The figure plots the quarterly frequency of the word "environment", calculated using a bagof-words approach (Loughran and McDonald, 2016), from 2000 to 2019. The blue line shows the frequency for our treatment group of 175 firms, composed of U.S. public oil gas firms (SIC 13) with non-missing corporate website information in COMPUSTAT. The black line plots the frequency for our 300 control firms, composed of randomly selected firms from the pharmaceutical industry (SIC 49). The green vertical line demarcates the incidence of the BP oil spill in Q2 2010. The dashed black vertical lines correspond to estimated structural breaks in the time series, with their associated confidence interval in red, as per the Bai and Perron (1998) procedure.



Table 1: Descriptive Statistics

The table presents the summary statistics for our website-based measure, four established measures of voluntary disclosure, and capital-market variables. Our website-based measure, Website Size, is defined as (the natural logarithm) of firms' website size. The four established measures of voluntary disclosure are: the number of management earnings forecasts (Management Forecast, period: 2000-2018); the number of disaggregated items in firms' financial statements (Disclosure Quality, period: 1979-2017); the word count of voluntary 8-Ks filings (Voluntary 8-K filings, period: 2005-2016); and the number of characters in the original 10-K filings (gross 10-K file size, period:1993-2017). The capital-market variables are: Bid-Ask Spread, defined as the quarterly median of the quoted bid-ask spread at the end of each trading day, calculated as the difference between the bid and the ask price divided by the midpoint price; Market Value, defined as the stock price multiplied by the number of shares outstanding at the end of each quarter (in \$US Millions); Share Turnover, defined as the quarterly mean of the daily turnover (the trading volume divided by the market capitalization at the end of each trading day); and *Return Variability*, defined as the standard deviation of daily stock returns over the quarter. All capital market variables and our website-based measure of disclosure are truncated at the 1st and 99th percentile. Panel A provides descriptive statistics for our website-based measure, established disclosure measures, and capital-market variables used in our analysis over the period 1997-2018. For each measure, we keep observations in the CRSP-COMPUSTAT merged database for which the market capitalization is non-missing. The Disclosure Quality measure is reported at an annual frequency while all other measures are reported at a quarterly frequency. Panel B compares the coverage of our website-based measure and the established measures relative to the full CRSP-COMPUSTAT universe of firms. We provide statistics from 2005, which is the year in which all four measures start to become available. We report the number of firms per measure and year and the corresponding proportion of the full CRSP-COMPUSTAT universe of firms. Panel C reports the distribution of market capitalization (at year-end) of firms covered by each disclosure measure. We provide statistics for the 2005-2016 period during which all measures are available. Panel D reports R-squares of regressions of the natural logarithm of our website-based measure on time, industry, and year fixed effects combinations. Panel E provides pairwise correlations for the natural logarithm of our website-based measure, the natural logarithm of (one plus) the established disclosure measures, and the natural logarithm of bid-ask spreads. All correlations in the matrix are significant at the 1% level (unreported for brevity).

	Count	Mean	Std. Dev.	P25	P50	P75
Disclosure variables:						
Website Size (Bytes)	209,186	11,928,648	$34,\!181,\!162$	124,944	$916,\!902$	6,511,753
ln(Website Size)	209,186	13.5	2.91	11.7	13.7	15.7
Management Forecast	121,486	3.50	2.98	1.00	3.00	4.00
Voluntary 8-K filings	184,425	7.41	2.32	7.31	8.00	8.57
Disclosure Quality	121,372	0.62	0.12	0.54	0.60	0.71
10-K file size (gross)	$357,\!071$	$3,\!111,\!371$	7,066,392	$129,\!161$	517,776	$2,\!438,\!572$
Canital market variables.						
Bid-Ask Spread (in %)	$184,\!594$	0.872	1.442	0.071	0.227	1.027
Market $Value_{t-4}$	$184,\!594$	2,513	$6,\!596$	113	456.9	1,755
Share $Turnover_{t-4}$	$184,\!594$	0.007	0.007	0.002	0.005	0.010
Return Variability $_{t-4}$	$184,\!594$	0.030	0.017	0.018	0.026	0.038

Panel A: Descriptive Statistics

Year	Full CRSP COMPUSTAT	Wayback Measure	%	Mgmt. Forecas	% t	8-K Filings	%	Disclos Quality	ure %	10-K Size	%
2005	4,924	2,674	54.3%	2,383	48.4%	$4,\!671$	94.9%	3,169	64.4%	4,703	95.5%
2006	4,839	2,713	56.1%	2,485	51.4%	4,593	94.9%	3,089	63.8%	4,614	95.4%
2007	4,776	2,851	59.7%	2,399	50.2%	4,584	96.0%	3,014	63.1%	4,542	95.1%
2008	4,535	2,866	63.2%	2,215	48.8%	4,352	96.0%	2,916	64.3%	4,291	94.6%
2009	4,248	2,729	64.2%	2,220	52.3%	4,118	96.9%	2,243	52.8%	4,071	95.8%
2010	4.076	2,773	68.0%	2,190	53.7%	3.964	97.3%	2,041	50.1%	3.906	95.8%
2011	3.890	2.940	75.6%	2.236	57.5%	3.806	97.8%	1.941	49.9%	3.757	96.6%
2012	3,777	3.006	79.6%	2,193	58.1%	3,701	98.0%	1.857	49.2%	3.648	96.6%
2013	3.741	3.074	82.2%	2.073	55.4%	3.680	98.4%	1.809	48.4%	3.630	97.0%
2014	3.840	3.236	84.3%	2.015	52.5%	3.782	98.5%	1.840	47.9%	3.719	96.8%
2015	3.848	3.354	87.2%	1.944	50.5%	3.810	99.0%	1.848	48.0%	3.699	96.1%
2016	3.734	3.373	90.3%	1.782	47.7%	3.717	99.5%	1.800	48.2%	3.547	95.0%
2017	3 684	3 396	92.2%	1 802	48.9%	N/A	N/A	1 608	43.6%	N/A	N/A
2018	3,649	3,415	93.6%	1,722	47.2%	N/A	N/A	N/A	N/A	N/A	N/A

Panel B: Coverage of Disclosure Measures

Panel C: Size Distribution Over 2005-2016

	Count	Mean	P10	P50	P90
ODOD COMDUCTAT II.	165 000	0 506	97	407	F 960
CRSP-COMPUSIAI Universe	165,908	2,526	37	407	5,869
Wayback Measures	$114,\!304$	$2,\!660$	43	519	$6,\!258$
Management Forecast	56,079	$3,\!857$	150	1,090	9,706
Voluntary 8-K filings	111,082	$2,\!649$	43	516	$6,\!253$
Disclosure Quality	$17,\!903$	2,075	39	441	4,760
Gross 10-K File Size	$108,\!679$	2,711	44	535	6,432

Panel D: Variance Decomposition of Website Size

	Sector (GICS)	Time	Firm
Sector (GICS)	0.042	-	-
Time	0.176	0.121	-
Firm	-	0.567	0.414

	Website Size	Management Forecast	Disclosure Quality	Voluntary 8-Ks	Gross 10-K File Size	ln(Bid-Ask Spread)
Website Size	1		·			
Management Forecast	0.120	1				
Disclosure Quality	0.297	0.232	1			
Voluntary 8-Ks	0.113	0.043	0.044	1		
Gross 10-K File Size	0.264	0.096	0.533	0.213	1	
$\ln(\text{Bid-Ask Spread})$	-0.352	-0.229	-0.394	-0.308	-0.506	1

Panel E: Correlation Matrix

ablished measures in firms' financial	lings (<i>Gross 10-K</i> s 1997-2018. The	er of observations	s and non-missing	om the table. For	ntheses below the	spectively.
<i>site Size</i> , on four est disaggregated items	l the size of 10-K fi num sample period	e measure. The numb	established measure	ression is omitted from	rs, reported in pare	5%, and $10%$ level, r
based measure, Web ccast); the number of	Disclosure $8-Ks$); and measure. The maxir	established disclosure	re and the respective	constant of the reg	y 10. Standard errc	nificance at the 1% , 5
thm of our website- s (Management Fore	filings (Voluntary L im of one plus the 1	e availability of the	vebsite-based measu	1. For brevity, the	sures are divided by	licate statistical sigr
f the natural logari it earnings forecasts	of voluntary 8-Ks the natural logarith	ons depending on th	erlap between our v	described in Table	l measures of disclo	l. ***, **, and * inc
f OLS regressions c mber of managemen	y); the word count e measure, we take	cross the specificati	bservations with ov	al-market variables	all four established	red at the firm leve
resents estimates o disclosure: the nu	(Disclosure Quality For each disclosure	aple period differs a	to the number of c	s for the four capit	the coefficients on	stimates, are cluste
The table p of voluntary	statements $File Size$).	effective san	corresponds	observation	readability,	coefficient e

Table 2: Website Size and Established Measures of Voluntary Disclosure

(12)	0.003**	(0.001) 156,208 0.570 Yes No Yes
(11)	0.041*** (0.003)	(0.002) 156,208 0.195 Yes Yes No
(10)	0.030***	(0.002) 156,208 0.136 Yes No No
(6)	0.005^{**} (0.002)	111,082 0.573 Yes No Yes
(8) ite Size)	0.058^{***} (0.004)	111,082 0.134 Yes No
e: $\ln(\text{Webs})$	0.057^{***} (0.004)	111,082 0.073 Yes No No
(6) ent Variabl	-0.030 (0.073)	29,057 0.562 Yes No Yes
(5) Depende	0.436^{***} (0.089)	29,057 0.211 Yes No
(4)	0.841^{***} (0.045)	29,057 0.172 Yes No No
(3)	0.007** (0.003)	76,882 0.552 Yes No Yes
(2)	0.041^{***} (0.005)	76,882 0.116 Yes No
(1)	0.051^{***} (0.006)	76,882 0.082 Yes No No
	Management Forecast Disclosure Quality Voluntary 8-Ks Gross 10-K File Size	Observations Adjusted R-square Year-Quarter FE Industry FE Firm FE

Table 3: Website Size and Liquidity

measure. In all specifications, we require observations to have non-missing information for the capital-market variables described in Table 1. For brevity, the constant of the regression is omitted from the table. For readability, the coefficient on ln(Website Size) is multiplied by 10. Standard The table presents estimates of OLS regressions of the natural logarithm of the bid-ask spreads on the natural logarithm of our website-based errors, reported in parentheses below the coefficient estimates, are clustered at the firm level. ***, **, and * indicate statistical significance at the $1\%,\ 5\%,\ and\ 10\%$ level, respectively.

	(1)	(2) Der	(3) endent Variable: ln(Bid-Ask Spre	(4) (4)	(5)
ln(Website Size)	-1.936^{***}	-0.639***	-0.088***	-0.055***	-0.065***
$\ln(\text{Market Value}_{t-4})$	(0.046)	(0.022) -0.392*** (0.007)	(0.018) -0.447*** (0.000)	(0.018) -0.451*** (0.000)	(0.015) -0.449*** (0.000)
$\ln(\text{Share Turnover}_{t-4})$		(0.007) -0.522*** (0.000)	(0.000) -0.433*** (0.007)	(0.000) -0.425*** (0.007)	(0.008) -0.250*** (0.008)
$\ln({\rm Return}~{\rm Variability}_{t-4})$		(0.003) 0.745^{***} (0.015)	0.506*** 0.506*** (0.014)	$\begin{array}{c} (0.001) \\ 0.507^{***} \\ (0.014) \end{array}$	(0.012) (0.012)
Observations Adiusted R-square	$184,594 \\ 0.124$	$184,594\\0.674$	184,594 0.808	$184,594 \\ 0.810$	184,594 0.874
Year-Quarter FE Industry FE Firm FE	No No	No No	Yes No No	Yes Yes No	Yes No Yes

Table 4: Text-Only Website Size and Liquidity

In all specifications, we require observations to have non-missing information on the capital-market variables described in Table 1. For brevity, the constant of the regression is omitted from the table. For readability, the coefficient on Website Size is multiplied by 10. Standard errors, reported The table presents estimates of OLS regressions of the natural logarithm of the bid-ask spreads on the natural logarithm of an adjusted website-based measure. In this table, our website-based measure, Size Website, is purged from non-text elements (i.e. images, videos, sounds, and applications). in parentheses below the coefficient estimates, are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

(1)	(6)	(3)	(4)	(4)
(+)		Dependent Variable: ln(Bid-Ask Spre.	ad) (*)	(0)
-2.067**	** -0.503***	-0.095***	-0.056**	-0.066***
(0.057)	(0.026)	(0.022)	(0.022)	(0.020)
	-0.394***	× -0.447***	-0.451^{***}	-0.449***
	(200.0)	(0.006)	(0.006)	(0.008)
	-0.532***	× -0.433***	-0.425***	-0.250^{***}
	(0.00)	(200.0)	(0.007)	(0.006)
	0.766^{***}	0.506***	0.507^{***}	0.222^{***}
	(0.016)	(0.014)	(0.014)	(0.012)
184,594	4 184,594	184,594	184,594	184,594
0.098	0.667	0.808	0.810	0.874
No	No	Yes	Yes	${ m Yes}$
No	No	No	Yes	No
No	No	No	No	Yes

r website-based measure, od in each panel depends of each measure. Panel ciations after controlling presents the conditional he measure. In Panel A specifications, we require number of observations with he coefficient on $Website$ im level. ***, **, and *	(5)	-0.064^{***} (0.015) -0.112^{***} (0.008)	Yes	$179,786 \\ 0.873 \\ \mathrm{Yes} \\ \mathrm{No} \\ \mathrm{Yes} \\ $
te natural logarithm of ou igh the actual sample peri details on the availability ents the conditional asso ry 8-K filings. Panel D logarithm of one plus t then it is missing. In all fable 1. As a result, the ich reports the total num table. For readability, tl s, are clustered at the fin	(4) Spread)	-0.037^{**} (0.018) -0.140^{***} (0.008)	Yes	$\begin{array}{c} 179,786\\ 0.810\\ \mathrm{Yes}\\ \mathrm{Yes}\\ \mathrm{No}\\ \mathrm{No} \end{array}$
the bid-ask spreads on th eriod is 1997-2018, althou railable. See Table 1 for of forecasts. Panel B preso for controlling for volunta intre, we take the natural slosure measure to zero w variables described in T ervations in Table 1, whi ssion is omitted from the r the coefficient estimates gement Forecasts	(3) ent Variable: ln(Bid-Ask	-0.042^{**} (0.017) -0.146^{***} (0.008)	Yes	$179,786 \\ 0.809 \\ Yes \\ No \\ No$
f the natural logarithm of lisclosures. Our sample po- easure of disclosures is av- trolling for management ditional associations afte For each disclosure meas <i>Quality</i>), we set the disc <i>Quality</i>), we set the disc from the number of obs from the number of obs from the number of the regres ted in parentheses below d 10% level, respectively. Panel A: Manag	(2) Depend	-0.455*** (0.020) -0.256*** (0.008)	Yes	179,786 0.701 No No
tes of OLS regressions o measures of voluntary o nich each established m al associations after con nel C presents the con ing for 10-K file size. Ind Panel B (<i>Disclosure</i> missing information fo ure of disclosure differs alization. For brevity, t Standard errors, repor ance at the 1%, 5%, an	(1)	-1.300*** (0.043) -0.888*** (0.017)	No	179,786 0.266 No No
The table presents estimat conditional on established on the time period for wh A presents the conditions for disclosure quality. Pa associations after controll (<i>Management Forecast</i>) a observations to have non- for each established meas non-missing market capit. <i>Size</i> is multiplied by 10. indicate statistical signific		ln(Website Size) Management Forecast	Controls	Observations Adjusted R-square Year-Quarter FE Industry FE Firm FF

 Table 5: Website Size and Liquidity Conditional on Established Disclosure Measures

			•		
	(1)	(2) Dependent V	(3) /ariable: Bid-Ask Spread	(4)	(5)
ln(Website Size)	-2.015***	-0.735***	-0.073***	-0.052***	-0.059***
Disclosure Quality	(0.048) 0.356^{***}	(0.025) 0.104^{***}	(0.019) -0.102***	(0.020)-0.035	(0.019) 0.178^{***}
	(0.061)	(0.030)	(0.026)	(0.032)	(0.034)
Controls	No	Yes	Yes	Yes	Yes
Observations	43,982	43,982	43,982	43,982	43,982
Adjusted R-square	0.135	0.662	0.812	0.813	0.881
Year-Quarter FE	No	No	Yes	${ m Yes}$	\mathbf{Yes}
Industry FE	No	No	No	${ m Yes}$	N_{O}
Firm FE	No	No	No	No	Yes

Panel B: Disclosure Quality

	(1)	(2) Deper	(3) ident Variable: Bid-Ask S _I	(4) pread	(5)
e Size) · 8-Ks	-1.365*** (0.051) -0.795*** (0.023)	-0.196*** (0.019) -0.113*** (0.009)	-0.043** (0.019) -0.091*** (0.009)	-0.017 (0.020) -0.103^{***} (0.009)	-0.036*** (0.013) -0.018*** (0.005)
	No	Yes	Yes	Yes	Yes
ons R-square rter FE FE	111,082 0.158 No No No	111,082 0.778 No No No	111,082 0.825 Yes No No	111,082 0.829 Yes No	$\begin{array}{c} 111,082\\ 0.916\\ \mathrm{Yes}\\ \mathrm{No}\\ \mathrm{Yes}\end{array}$

8-Ks
Disclosure
Voluntary
ö
Panel

	(5)	-0.069^{***} (0.016) -0.015^{***}	(0.004)	Yes	156,208	0.882	$\mathbf{Y}_{\mathbf{es}}$	No	Yes
Panel D: Gross 10-K File Size	(4) pread	-0.055^{***} (0.018) 0.007	(0.005)	Yes	156,208	0.816	Yes	\mathbf{Yes}	No
	(3) lent Variable: Bid-Ask S _I	-0.082^{***} (0.018) 0.017^{***}	(0.006)	Yes	156,208	0.815	Yes	No	No
	(2) Depend	-0.421^{***} (0.020) -0.237^{***}	(0.004)	Yes	156,208	0.723	No	No	No
	(1)	-1.288*** (0.046) -0.404^{***}	(0.006)	No	156,208	0.308	N_{O}	No	No
		ln(Website Size) Gross File Size		Controls	Observations	Adjusted R-square	Year-Quarter FE	Industry FE	Firm FE

10-K File Si ç Ċ ol D.

Table 6: International Coverage

The table presents the availability of our website-based measure for all countries (excluding the U.S.) with at least 10 unique listed firms in 2015 in the COMPUSTAT Global database. For each country, we list the total number of firms in COMPUSTAT Global in 2015, the number of firms with website information in the Wayback Machine data, and the coverage percentage (Wayback Machine/COMPUSTAT Global).

Country	Unique	Wayback	%	Country	Unique	Wayback	: %
	\mathbf{Firms}	Firms			\mathbf{Firms}	Firms	
China	4,309	4,234	0.98	Peru	108	102	0.94
India	$3,\!955$	$3,\!831$	0.97	Oman	104	101	0.97
Japan	$3,\!254$	3,238	0.99	Croatia	91	89	0.98
Canada	$2,\!186$	1,937	0.89	Cyprus	90	78	0.87
Taiwan	$1,\!991$	1,978	0.99	Austria	86	86	1.00
Australia	$1,\!980$	1,905	0.96	Bulgaria	84	83	0.99
United Kingdom	$1,\!959$	1,888	0.96	Argentina	80	74	0.92
Hong Kong	$1,\!489$	$1,\!442$	0.97	Morocco	78	74	0.95
Korea	$1,\!428$	$1,\!414$	0.99	Cayman Islands	78	75	0.96
Malaysia	984	974	0.99	Tunisia	77	69	0.90
Poland	852	845	0.99	Ireland	68	68	1.00
Germany	772	759	0.98	Bermuda	60	52	0.87
Thailand	758	741	0.98	Mauritius	58	54	0.93
Sweden	753	748	0.99	Portugal	53	53	1.00
France	712	703	0.99	Kenya	53	52	0.98
Singapore	687	653	0.95	Colombia	52	52	1.00
Indonesia	605	595	0.98	Luxembourg	52	50	0.96
Israel	495	419	0.85	Qatar	45	45	1.00
Vietnam	451	451	1.00	Ukraine	41	39	0.95
Pakistan	441	438	0.99	Zimbabwe	38	36	0.95
Turkey	418	414	0.99	Hungary	37	37	1.00
Brazil	390	361	0.93	Jamaica	37	35	0.94
Italy	365	357	0.98	Bahrain	37	36	0.97
South Africa	327	325	0.99	Palestine	34	33	0.97
Switzerland	280	276	0.99	Lithuania	34	33	0.97
Sri Lanka	277	239	0.86	Kazakhstan	32	32	1.00
Russian Federation	274	269	0.98	Slovenia	31	31	1.00
Philippines	258	250	0.97	Latvia	27	26	0.96
Bangladesh	251	250	0.99	Malta	27	27	1.00
Norway	242	239	0.99	Iceland	25	25	1.00
Jordan	216	171	0.79	Ivory Coast	25	23	0.92
Greece	209	205	0.98	Ghana	24	23	0.96
Chile	206	185	0.90	Serbia	24	22	0.92
Egypt	206	200	0.97	Zambia	22	20	0.91
Spain	205	197	0.96	Botswana	21	19	0.90
Kuwait	195	187	0.96	Trinidad and Tobago	20	20	1.00
Saudi Arabia	189	188	0.99	Czechia	17	17	1.00
Finland	173	172	0.99	Estonia	16	16	1.00
Denmark	172	168	0.98	Venezuela	14	12	0.86
New Zealand	166	159	0.96	Tanzania	14	13	0.93
Mexico	149	147	0.99	Slovakia	14	14	1.00
Netherlands	148	144	0.97	Virgin Islands	13	10	0.77
Nigeria	147	142	0.97	Guernsey	13	12	0.92
Belgium	143	138	0.97	Jersey	12	12	1.00
Romania	131	128	0.98	Lebanon	10	9	0.90
United Arab Emirates	128	124	0.97				

Appendix A: Examples

We present screenshots from Q4 of 2017 of the main landing corporate webpages of two pharmaceutical firms, *Celcuity Inc.* and *Teligent Inc.*, included in our sample.



Appendix B: List of Keywords

The table describes the list of keywords used to assign content categories to each URL. To reduce misclassification issues, keywords are not reduced to their root form (i.e. *stemmed*) before they are manually classified. This results in several inflections for each word (e.g. plural) being contained in the list below.

Topic	Keywords
Product, Strategy and	buy, client, clients, consumer, contract, customer, customers, partner, partners,
Processes	partnership, supply, vendor, partnerships, supplier, suppliers, competition, ads, advertising, brand, brands, launch, launches, marketing, portfolio, price, pricing, product, products, project, purchase, quality, retail, service, services, solution, solutions, campaign, certification, certified, demand, license, licensing, optimization, prices, delivery, building, capacity, data, distribution, energy, equipment, industrial, industry, invest, investment, manufacturing, operations, platform, process, processing, production, protection, container, device, devices, electric, electronic, electronics, engineer, facilities, gas, industries, logistics, machine, oil, operating, operation, pipeline, plant, platforms, platinum, steel, storage, transport, truck, utilities, innovation, research, resource, resources, science, tech, technical, technologies, technology, expert, expertise, ideas, innovations, innovative, intelligence, knowledge, patent, projects, scientific, solar, business, culture, expands, market, markets, mission, shop, strategic, strategy, success, value, values, vision, expand, expansion, opportunity, strategies, drug, stores, store
Investor and Corporate Governance	account, asset, assets, audit, benefits, book, cash, sale, sales, accounting, accounts, balance, benefit, billing, books, costs, board, director, directors, employee, employees, executive, governance, head, officer, officers, president, senior, chairman, announce, announces, calendar, earnings, filings, performance, quarter, release, releases, report, reporting, reports, results, schedule, dividend, acquire, acquisition, agreement, analyst, analysts, bank, banking, capital, commercial, companies, company, compliance, corporate, cost, coverage, equity, exchange, finance, financial, financing, growth, insurance, investor, investors, legal, loan, meeting, money, offer, offering, savings, sec, shareholder, statement, stock, target, tax, acquires, announcement, budget, committee, compensation, deal, affiliate, closing, corporation, disclosure, executives, fiscal, fraud, fund, income, investments, law, listing, loans, merger, outlook, ownership, rate, rates, rating, regulatory, result, revenue, shareholders, statements, trading, transaction
Geography	africa, america, american, americas, angeles, arizona, asia, atlanta, austin, australia, boston, brazil, california, canada, canadian, carolina, chicago, china, colorado, columbia, country, county, diego, east, europe, european, florida, france, francisco, french, georgia, german, germany, illinois, india, indiana, italian, italy, japan, jersey, kansas, location, locations, london, mexico, miami, michigan, office, ohio, pacific, paris, philadelphia, santa, seattle, singapore, southern, spain, spanish, state, texas, usa, vegas, virginia, washington, west, western, world, worldwide, york, border, branch, houston, dallas, denver, phoenix, coast, offices, regional
Human Resources	career, careers, college, education, employment, experience, job, jobs, opportunities, recruiting, school, students, training, university, campus, student