European Accounting Review

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rear20

Mobilizing Text As Data

Jihun Bae, Chung Yu Hung & Laurence van Lent

To cite this article: Jihun Bae, Chung Yu Hung & Laurence van Lent (2023) Mobilizing Text As Data, European Accounting Review, 32:5, 1085-1106, DOI: 10.1080/09638180.2023.2218423

To link to this article: https://doi.org/10.1080/09638180.2023.2218423

European Accounting Review, 2023
Vol. 32, No. 5, 1085–1106, https://doi.org/10.1080/09638180.2023.2218423

Check for updates

Mobilizing Text As Data

JIHUN BAE [a], CHUNG YU HUNG [b] and LAURENCE VAN LENT [c]

[a] *Erasmus University Rotterdam, Rotterdam, The Netherlands;* [b] *University of Melbourne, Melbourne, Australia;*
[c] *Frankfurt School of Finance and Management, Frankfurt am Main, Germany*

(*Received: 12 December 2022; accepted: 18 May 2023*)

ABSTRACT Textual analysis methods have become increasingly popular and powerful tools for researchers in finance and accounting to extract meaningful information from unstructured text data. This paper surveys the recent applications of these methods in various domains, such as corporate disclosures, earnings calls, investor relations, and social media. It also discusses the advantages and challenges of different textual analysis methods, such as keyword lists, pattern-based sequence classification, word embedding, and other large language models. We provide guidance on how to choose appropriate methods, validate text-based measures, and report text-based evidence effectively. We conclude by suggesting some promising directions for future research using text as data.

Keywords: Textual analysis; Finance; Accounting; Text data; Natural language processing

1. Introduction

One may question why we choose to contribute to the discussion on the use of text as data in economics, finance, and accounting, given the presence of several authoritative views (Bochkay et al., 2023; Gentzkow et al., 2019; Loughran & McDonald, 2011). Those seeking technical insights on natural language processing methods, their application pitfalls in capital markets or managerial research, or the latest advances in language models will find ample satisfaction in these prior works.[1] Our article pursues three objectives from a different perspective. Primarily, we aim to inspire innovative research questions using these developments. Secondly, we discuss various text sources that (accounting) researchers can explore to answer significant questions. Lastly, we offer guidance on presenting text-based evidence to establish its empirical credibility. We target a broader set of researchers interested in leveraging text data for innovative research. We explore opportunities for accounting researchers to broaden their work's impact using text data and illustrate how accounting expertise can contribute to research topics where the accounting angle might not be immediately apparent.

Numerous societal issues, such as social justice, environmental concerns, political discord, democratic values' erosion, and public health, provide opportunities for accounting researchers

**Correspondence Address*: Chung Yu Hung, University of Melbourne, 198 Berkeley Street, Melbourne, Victoria 3053, Australia. Email: chung-yu.hung@unimelb.edu.au
Paper accepted by Beatriz García Osma.

[1] Ranta and Ylinen (2022) discuss machine learning in management accounting research. European Accounting Review, vol. 29, Issue 1 offers a special section devoted to textual analysis studies in accounting. Mahlendorf et al. (2022) provide further stimulating ideas about using textual data in management accounting research.

to contribute. Mainstream accounting research has been hesitant to engage with these issues, often considering them outside the narrow scope of accounting research. This reluctance is unfortunate, as understanding these problems, a prerequisite for their resolution, often involves *measurement*. Addressing issues like gender pay disparity, global warming, and racial inequalities requires adequate measures to express their impact and assess the problem's magnitude. Accounting research can excel here, particularly when examining how these issues affect firms or how firms impact society. Accountants are experts in quantifying firms' operations and engagement with the outside world. They deeply understand information production and consumption within firms and markets and have thoroughly considered verification, i.e. enhancing information credibility. This understanding can readily be applied to address broader societal issues.

Addressing these significant questions with an accountant's toolkit fulfills the criteria for successful empirical work. The question remains: why should accountants be interested in using text as data? An irreverent answer might be: why not? As Gentzkow et al. (2019) note, new technologies have made vast quantities of digital text available. This trend is expected to continue, if not accelerate. The enormous volume and diversity of currently accessible text provide not only large data sets (in both cross-sectional and time series senses) that previous researchers could only dream of but also valuable information about complex, hard-to-measure phenomena. Textual data can often be explored for context and offer multiple key informants' perspectives. We further develop these themes in this article.[2] [3]

Accountants possess at least two sustainable comparative advantages when leveraging text as data. First, as a profession, we have accrued centuries of expertise in measuring complex, varied transactions in firms and translating these transactions into meaningful statistics (Bierman, 1963; Homburger, 1961). *Measurement* is ingrained in our discipline. Accounting students have been trained to discern the nature of transactions and recognize their economic substance. Although our attention has been selective–focusing on specific transactions and mapping them into financial/monetary terms (Hines, 1988)–our expertise in developing a sophisticated set of rules, procedures, and approaches to deal with the complexity of understanding business dealings remains valuable (Andreicovici et al., 2020).[4]

Our second professional advantage is our training to consider information from the *broad* conceptual perspective of 'usefulness.' This framework entails thinking about information objectives (which may vary across constituents), users (which could be numerous), and mapping the properties of information to those specific users' information needs. Such a framework is adaptable to consider various information sources, users, uses, and properties. As a result, we have studied firm reporting in different forms, observed its expansion over the past 50 years, and gained insights into how firms communicate with stakeholders. Much of this communication is captured in text (although intriguing non-text, non-financial material also exists). We are as much text experts as we are numbers experts, giving us an advantage over other social scientists.

---

[2]We acknowledge the tradition of (textual) content analysis in accounting research, dating back to studies of annual report readability in the 1950s and 60s (Poshalian & Crissy, 1952; Soper & Dolphin, 1964). Content analysis has also been used to learn about management attitudes towards social responsibility and (environmental) performance (Bowman & Haire, 1976; Chan, 1979; Neimark, 1983) and to study the correlation between narrative disclosures and financial reporting (Frazier et al., 1984; Ingram & Frazier, 1980). Perhaps most famously, positive accounting theory originated in the content analysis of comment letters to the FASB (Watts, 1977).

[3]See also, Jones and Shoemaker (1994).

[4]Perhaps our most remarkable insight as accountants is that measurement is never perfect. To make measurement worthwhile, judgment needs to be involved (Bierman, 1963). Consequently, reported profits are an opinion, not a fact (Klaassen, 1992). When contributing to measurement problems outside the financial reporting domain, the objective should be charting these imperfections rather than pretending they do not exist.

Our comparative advantage also leads us to be cautious about calls for accounting researchers to become more proficient in advanced machine learning techniques (Bochkay et al., 2023). In fact, we advocate simplicity as the guiding principle in text-based work. The rapid methodological advances in natural language processing can easily tempt researchers to apply new, sophisticated tools. It is crucial to start with an important research question rather than a method to avoid succumbing to this temptation. Simple methods help convince readers, yield economically meaningful interpretations, and are less prone to researcher choices/interventions. Given this preference, we limit our discussions to supervised/semi-supervised methods and exclude 'let-data-speak-for-themselves' methods (e.g., unsupervised methods), as their interpretation is often less straightforward (and has been described as 'art' rather than science). Using various examples, we illustrate essential principles when developing text-based measures. Our discussions focus on three aspects. First, we discuss a specific method, pattern-based sequence classification, to demonstrate the advantages of simplicity in text-based research. Second, we argue that validating a text-based measure is essential for building empirical credibility. Third, we discuss strategies for showing that a newly developed metric helps answer economically significant questions. However, in the next section, we begin our discussion with a tour of potential text sources–both familiar and perhaps less so.

2. Text Data-sources

Text data appear in various forms such as newspaper articles, social media posts, research reports, legal documents, and marketing materials. Utilizing these data, accountants can draw upon their conceptual framework, recognizing two primary financial reporting objectives: *valuation*, which provides information for users to assess a firm's future cash flow prospects, and *stewardship*, which offers information to evaluate management's handling of the firm's resources. Numerous desirable characteristics stem from these objectives, including relevance (and materiality) and faithful representation (and completeness, neutrality, and freedom from bias). These concepts assist in analyzing textual disclosures of and about the firm beyond its financial statements. For example, a CFO's tweets may predict future cash flows, while court documents can reveal previously unknown management performance details. The wide range of available textual records provides new insights into management's stewardship and allows for further tests accounting for users' more comprehensive information sets when predicting future performance. Researchers can examine user responses to company communication, information transmission among market participants, and firms' credibility-enhancing interventions.

This section emphasizes text data sources and their potential unconventional uses. Instead of providing structure to these data sources, we note that considering ways to characterize text can refine research designs. For instance, distinguishing between mandatory and voluntary text documents relates to susceptibility to strategic disclosure incentives. Likewise, scripted versus spontaneous speech likely experiences differing degrees of management's ability to present select facts. It is also useful to differentiate between firm-initiated text, stakeholder-initiated text, and mixed forms reflecting firm-user interactions (such as investor day discussion transcripts). Depending on the specific research question using text data, these attributes may render a source more or less suitable.

2.1. Firm Disclosures

Macroeconomists have started recognizing the value of regulatory filings and financial statements as a treasure trove of valuable microdata–something accounting researchers have long known. A rich variety of accounting-related text is available, particularly outside the US, where corporate reporting typically involves less structure and greater heterogeneity. This is especially true in Europe, where management enjoys significant discretion over financial reporting package contents, organization, and presentation and layout methods (International Accounting Standards Board, 2017). Recent work using text data from these sources demonstrates that firms' attention to the macroeconomy varies with the business cycle (Flynn & Sastry, 2022) or shows that global capital flows depend on executives' country risk perceptions (Hassan et al., 2023b). Examples like these likely mark the beginning of numerous explorations.

Accounting researchers might yield to economists, finance scholars, and management experts interested in using these data to address open questions in other fields. Alternatively, they might actively engage with these questions and collaborate with researchers outside accounting. Doing so successfully necessitates somewhat of a mindset change. Traditional primary concerns for accountants, such as (earnings) manipulation and strategic disclosure incentives, are likely secondary problems for those seeking to harvest microdata about firms to learn about individual economic agents' perceptions and actions that are otherwise difficult to obtain.[5] Our advanced understanding of disclosures (and the incentives of those involved) contributes to the value we provide in discourses that rely on data from these documents. Indeed, properly accounting for strategic disclosure incentives might help assess how reliable and valuable disclosed information can be to specific decision-makers. The dark side of this understanding is that we might be too careful and reluctant to use the data beyond the usual applications. Having documented that executives have strategic incentives to disclose and withhold information might lead one to conclude erroneously that nothing factual can be learned by studying the text. Then, accounting research becomes more of an impediment than an expedient to the scientific effort. It would also be a misreading of the large body of evidence in accounting that shows text to be incrementally useful on average.

2.2. *Conference Call Transcripts*

Conference calls are awesome! Few alternative text-as-data sources are publicly available for a global sample of firms, allowing insights into senior managers' thoughts. Moreover, conference calls take place quarterly, enabling researchers to access this information rapidly.[6] Using conference calls is akin to conducting a survey (Graham et al., 2005), but with minimal cost and on historical and real-time data. Though researchers may not ask the questions they desire, call participants are effective interlocutors.[7]

Conference calls, usually voluntary and held quarterly, begin with management presentations, followed by Q&A sessions with market participants, which are less controlled by corporate management (Matsumoto et al., 2011).

Viewed as information marketplaces, conference calls let firms disclose desired information and respond to market participants' inquiries. In contrast, regulatory filings with pre-determined

---

[5]The point we are making here is not that earnings manipulation and strategic disclosures are unworthy of study. We gently suggest that in allocating our profession's resources, it might be more profitable to widen our horizon. Rather than writing another paper about managers attempting to manipulate earnings to fool investors, we do something different.

[6]A growing body of work in accounting focuses on earnings call transcripts, analyzing language features to detect misconduct, reveal undisclosed news (Chen et al., 2018; Hobson et al., 2012; Hollander et al., 2010; Larcker & Zakolyukina, 2012; Lee, 2016), investigate information processing (Allee & DeAngelis, 2015; Bochkay et al., 2020; Bushee et al., 2018; Bushee & Huang, 2021), and explore managers' attributes and communication behaviors (Bochkay et al., 2019; Brochet et al., 2015, 2019; Davis et al., 2015).

[7]We acknowledge the accounting evidence showing that managers use language and actions to present the firm or their behavior in a positive light (Mayew, 2008; Mayew et al., 2013); we discuss strategies for addressing strategic disclosure in later sections (e.g., Section 4.2).

formats exhibit limited year-over-year variation (Brown & Tucker, 2011; Cohen et al., 2020). Market participants pressure management for information, reducing the likelihood of boilerplate disclosures. Instead of being strategically dishonest, management may be genuinely uncertain about the information demanded by market participants (Suijs, 2007). Conference calls offer flexibility in providing information in response to analysts' questions when investors' needs are unclear (Frankel et al., 1999). Compared to regulatory filings, conference calls can disaggregate the information, often discussing soft information (Campbell et al., 2019), judgments, and perceptions. Direct information from economic agents is rare in other data sources.

Besides earnings calls, another valuable source of corporate transcripts is investor day meetings. These events have received little attention from researchers (except for Kirk & Markov, 2016), but they deserve more exploration. Investor day meetings are longer (1-1.5 days) and more detailed than earnings calls, and they involve a wider and more diverse range of senior managers. This allows researchers to gain insights into the communication dynamics within the firm that are hard to observe otherwise (Bochkay et al., 2019; Impink et al., 2021; Li et al., 2014).

In addition to textual analysis, conference calls offer opportunities to apply machine-learning algorithms to visual materials (e.g., PowerPoint slides) accompanying management's opening statements.[8] These documents, not captured in transcripts, largely await further analysis.

2.3. *Job Openings*

Firms' job postings for new personnel are a relatively under-explored text source that offers insights into their internal operations (Darendeli et al., 2022). Besides basic details like the open position, salary range, and required skills, these postings can reveal aspects of company culture, job location, and executive decision-making. Recent studies have begun to explore these possibilities, such as Kalmenovitz (2019) and Wu (2020), which use the Burning Glass database to infer regulatory burden based on compliance personnel hiring. Cai et al. (2021) measures firms' human capital in crisis management through job postings requiring relevant skills, while Acemoglu et al. (2022) examines AI-related job vacancies' effects on labor markets. The Burning Glass database has extensive coverage, particularly in EU countries.

The full potential of such datasets likely lies in creatively combining multiple text sources. An intriguing example is provided by Bloom et al. (2021), who use job openings, patent registries, and conference call transcripts to study the spread of disruptive technologies in the economy. They employ textual analysis to identify technological breakthroughs in patent filings, find exposed firms through conference calls, and track the dispersion of knowledge and new skills via technology job hiring.

2.4. *Social Media Postings*

In addition to conference calls, firms increasingly interact with stakeholders through social media, generating a new text data source to examine their dealings with investors, interest groups, consumers, politicians, and others. Social media data is attractive due to its timeliness, as near real-time monitoring allows researchers to capture immediate responses to major events or reconstruct detailed timelines of firms' reactions to shocks.

For instance, Fritsch et al. (2022) analyze tweets related to a climate change crisis to examine firms' disclosure activities on Twitter, while Fritsch (2022) investigates whether managers can improve their internal forecasts by considering social media users' feedback. Social media text,

[8] Pattern-based sequencing algorithms can be utilized to analyze other unstructured data, such as executives' body language or vocal tones.

characterized by its brevity, typos, non-sequiturs, and non-verbal expressions, poses challenges for standard computational linguistics algorithms but can be tackled with dedicated procedures (Liu et al., 2021).

Verbal employee feedback on Glassdoor can complement the numerical assessments used by researchers like Huang et al. (2020) to study how employees anticipate future performance. Moreover, textual research can leverage investor discussions on platforms such as Seeking Alpha and Motley Fool, as well as financial analysts' views on Estimize (Jame et al., 2016).

3. Developing Text Measures

We focus on three simple methods for creating text measures that require minimal researcher input to select the words to count. These methods are appealing because they are easy to comprehend, have clear economic meanings, and have been successfully used in top journals in economics, finance, and accounting. We do not claim that these are the only or the best methods for every situation.[9]

Word counts are appealing because they have an intuitive (economic) interpretation. For example, for a conference call, we can divide the word count by the transcript length to get the share of conversation between analysts and management on a given topic (as captured by the relevant words) (Hassan et al., 2019). The challenge is to identify these relevant words. For instance, if we want to know whether firms discuss social movements like 'black lives matter' in their disclosures, we need to find the words (or word combinations) that indicate racial inequality in a document. This is the key task in creating a text-based measure based on word counts.[10] Word counts can also capture more subtle concepts by using 'conditional word counts.' For example, Hassan et al. (2019, 2023a) and Hassan et al. (2023b) are interested in measuring the risk and sentiment related to an economic shock (e.g., a political shock like Brexit, Trump's election, or the pandemic). They do this by counting words that are modified by words indicating risk and sentiment in addition to the words pertaining to economic shocks. This flexible approach can also capture other information attributes such as timeliness or bias.

Keyword Lists. At times, researchers have it easy. For topics such as Brexit, the Fukuyama nuclear disaster, or the Covid pandemic (Hassan et al., 2023a, 2023b), pertinent keywords are evident and can be compiled without much controversy. Jamilov et al. (2021) utilize think tank documents and studies to gather terms related to cybersecurity. This method, which relies on reputable policy institutes or third-party research centers, is sensible but has drawbacks. Authoritative texts may contain language common among (technical) experts but might not reflect how analysts and management discuss these issues, leading to imprecise measurements due to language misalignment.

A more difficult scenario arises when the topic of interest is broader and cannot be captured by a few keywords. For instance, when investigating how firms are affected by political events, *prespecifying* all political topics (e.g., healthcare, presidential elections, state-level political

[9] Other methods, such as unsupervised methods, can find latent patterns of texts without predefined topics. However, these methods may produce topics that are hard to interpret and prone to post-hoc justification by researchers. See Bochkay et al. (2023) for more details about unsupervised methods.

[10] Word counts may miss some context and nuances of meaning, but they avoid the main problem that plagued early content analysis work in accounting and limited its impact (Jones & Shoemaker, 1994). That problem is relying on human coders to interpret the *meaning* of a sentence, which assumes that they can consistently recognize homographs, symbolic meaning, and the meaning of words over time. Computational linguistics has advanced to allow computers to learn the meaning of a word from its context, which gives researchers more tools without resorting to manual content analysis and its potential measurement error.

events, or international relations) and listing related word combinations is daunting and likely impossible.

Fortunately, the language patterns (i.e., word sequences) used to discuss political topics are likely to differ from those on accounting standards, science, or technology. Pattern-based sequence classification can differentiate language patterns among various topics, allowing researchers to recognize patterns specific to political topics but not others. This method is particularly beneficial for developing text-based measures without unique word combinations (e.g., Brexit or COVID-19). We discuss pattern-based sequence classification in detail before addressing semi-supervised approaches for identifying keyword lists.

3.1. Pattern-Based Sequence Classification

Consider an archaeologist who uncovers a ceramic fragment with engraved symbols, determined to originate from one of three languages: Linear A, Linear B, or Cretan Hieroglyphic. The symbols form an ordered list, representing a sequence or pattern. Pattern-based sequence classification identifies which of the three language classes the symbol combination is most common in, enabling the labeling of the fragment.

Pattern-based sequence classification is a straightforward yet powerful method for classifying patterns into specific classes.[11] In our context, examples typically involve *text* patterns (word combinations or n-grams) assigned to a 'language' class (e.g., a particular topic).[12] For instance, we may want to determine whether a firm discusses cybersecurity in its 10-K filing or if certain analysts consistently ask questions about gender pay gaps and other inequality issues in a company. The corresponding 'language' classes are binary (cybersecurity or not-cybersecurity; gender pay gap or no gender pay gap), and we aim to identify the word combinations unique to each class. However, pattern-based sequencing has broader applications than word sequences, such as classifying musical notation or identifying the origin of folk music tunes (van Zaanen & Kanters, 2015). It can also analyze patterns in body language in videos, differentiating between emphatic or non-emphatic and authoritative or non-authoritative gestures. Viewing the world as a collection of classifiable patterns to derive meaning provides numerous opportunities for new research.

To determine the boundaries of a language class, we follow a two-stage process: training and classification. First, we select unambiguous training data and analyze all potential patterns it contains. The crucial step is identifying which patterns best aid classification. Classification involves comparing a new fragment to patterns from the training data to determine the best match for the class.

In the training stage, patterns are extracted from the data, with the training set divided into fixed-order word combinations. Each pattern is assigned a weight per class, reflecting its classification strength. A common approach employs 'term frequency*inverse document frequency (tf*idf)' to determine this strength. Originally from information retrieval, tf*idf gauges document relevance within a large collection given a search term. It consists of term frequency (tf) and inverse document frequency (idf). Tf measures a pattern's regularity, while idf assesses its discriminative power. Tf counts a pattern's occurrences in a document and normalizes the result by document length to avoid bias toward longer documents. High text frequency patterns are more

---

[11]This section draws heavily from the work of our former colleague Menno van Zaanen and our discussions with him over the years.

[12]The pattern-based sequence classification framework can be applied to various data types, such as video or music notation. In the context of text data, 'pattern' refers to a list of ordered word combinations (e.g., bigrams), while 'class' refers to the topic defined by the construct of interest, such as cybersecurity and non-cybersecurity or Brexit and non-Brexit topics.

helpful for classification, as they better describe a specific class. However, a pattern appearing in all classes is not useful for determining class membership. Patterns unique to a class are valuable for classification, as their presence indicates class affiliation. Idf captures a pattern's uniqueness to a class and is computed as the logarithm of the ratio of total classes in the training data to the number of classes containing the pattern.[13]

For each pattern in the training data, we calculate the tf*idf for each class. Patterns that appear in all classes have zero idf and zero weight, regardless of their tf. This makes sense because these patterns cannot distinguish between classes. Patterns that appear frequently in one class have higher weights to reflect their relevance. Some patterns with high tf*idf values are clearly related to the language class we are interested in, but some are less obvious. For example, among the top 120 political bigrams in Hassan et al. (2019), 'President Obama' is an evident political bigram, but 'support for' and 'shall have' are more general and seem not specific to political language. These general bigrams with high classification strength do not undermine this method but show how different language classes use different words, word orders, and sentence structures. Political language has different phrases than the language used to describe corporate financial performance.

In classification, an unseen document's class is determined by extracting all patterns, computing the tf*idf score for each matched pattern from the training set to the document, and considering each class. Each pattern will have as many tf*idf values as there are classes. Summing the tf*idf values over the matched patterns provides a document score for that class, with multiple scores corresponding to different classes. Classification involves selecting the class with the highest summed tf*idf score.

Pattern length matters, as short patterns frequently occur, resulting in high tf and low idf values and lower distinguishability. Longer patterns are less common and have low tf and high idf values, making them more distinguishable. However, longer patterns are not necessarily preferred. As patterns lengthen, it becomes harder to find them during the classification stage, reducing the likelihood of identifying the pattern that unambiguously indicates a new document's class. Empirically, using bigrams over unigrams significantly improves performance in finance and accounting applications (Hassan et al. 2023; Sautner et al. 2023a, 2023b) while even longer patterns seem less effective (van Zaanen & Kanters, 2015).

An example can illustrate the idea better. Consider these three patterns: 'Obama care,' 'corporation tax,' and 'depreciation expenses.' These patterns were found in the training stage from a data set of political and business texts to separate political from financial language (two language classes in this case). In the training stage, we see that the first bigram, Obama care, is common in the political texts but *not* in the financial texts (high tf for politics class but zero tf for financial class). The second bigram, corporate tax, is common in both political and financial texts (high tf for both politics and financial classes). The third bigram, depreciation expenses, is absent in the political texts but frequent in the financial texts (zero tf for politics class but high tf for financial class).

'Obama care' and 'depreciation expenses' have high idf values because they are unique to one of the two classes, but 'corporate tax' has zero idf value because it appears in both politics and financial classes. We summarize the three examples in Table 1.

The tf*idf value in a given class shows how useful a pattern/bigram is to classify a new document in that class. Based on the above discussion, 'Obama care' has the highest tf*idf value in the political text class, and the other two bigrams have zero tf*idf values. Obama care is more

[13] For more on tf*idf, see Loughran and McDonald (2016). For applications, see Engle et al. (2020), Hassan et al. (2019), and Sautner et al. (2023a).

Table 1. Illustration of tf*idf weights.

Patterns (bigrams)	Term frequency in political class (tf)	Inverse document frequency (idf)	tf*idf in political class
Obama care	High	High ($\log \frac{2}{1}$)	High
Corporation tax	High	Zero ($\log \frac{2}{2}$)	Zero
Depreciation expenses	Zero	High ($\log \frac{2}{1}$)	Zero

likely a political term than ‘corporate tax’ and ‘depreciation expenses.’ The bigram ‘depreciation expenses’ is clearly not a political term because it never appears in the political text (zero tf). The bigram ‘corporate tax’ often appears in political text, but it also appears in the nonpolitical text (zero idf). If a new document that we want to classify has the phrase ‘Obama care,’ we can confidently label it as a political text.

Thus far, the crucial step of selecting the training data has been overlooked. The aim is to identify patterns that differentiate the classes of interest. Accordingly, the training data should include archetypal language for each class. For instance, Hassan et al. (2019) differentiates political discourse in earnings conference calls from other firm-related discussions. The paper employs two ‘training libraries’ containing newspaper articles on domestic politics, an American politics textbook, articles on business topics, and a financial accounting textbook. Later, they broaden their scope, using online resources to distinguish among various American political topics, refining the classification task from two categories (politics or not) to eight separate political topics.

Textbooks and newspaper articles offer valuable training library material, as they are often labeled by class, such as ‘domestic politics’ and ‘American politics.’ However, the language in textbooks differs from the conversational style of conference calls, necessitating adjustments. The Santa Barbara Corpus of Spoken American English can be used for this purpose.

Labeled training sets are essential for pattern-based sequence classification, offering the advantage of requiring minimal researcher intervention, limited to selecting library texts, which can be traceable and subject to robustness checks.

One can think of corporate disclosures as having built-in labeled data for training purposes. Accounting researchers can utilize domain-specific labels, such as market-based tags (e.g., stock price reactions) and management-assigned labels like section headers in annual reports, sustainability reports, or modern slavery reports. Section headers facilitate the creation of ‘clean’ corpora on specific topics, such as risk, remuneration, strategy, business model, and workforce. Although management-assigned labels may introduce bias, an average sample of such text is likely more representative of financial discourse than an exogenous ‘authoritative’ text. Regulatory developments regarding tagging taxonomies are expected to further enhance the appeal of accounting disclosures as source material.

Text classification is a preliminary step in constructing an entity-level measure (Bloom et al., 2021; Hassan et al., 2019). The algorithm generates a tf*idf score for each class and every word combination. If there are only two classes, the pattern will have either a positive or zero score. The source document can then be used to sum all non-zero tf*idf scores to create a measure. Alternatively, matched pattern occurrences can be counted, generating a measure unweighted by tf*idf.[14]

Once the class-specific keywords are identified, they can be further adapted for specific objectives. For example, Hassan et al. (2019) use a conditional count to capture political risk, considering the frequency of specific word combinations related to politics and risk or

[14] Hassan et al. (2019) find this choice to be of little consequence in their context.

uncertainty synonyms. Researchers investigating deceptive disclosures in climate and carbon emissions contexts can adopt a similar approach. By identifying word combinations indicative of climate change (using lists from sources like Sautner et al. (2023a) or keywords from Engle et al. (2020)), a conditional count based on the presence of 'weasel words' near climate change language can be constructed. Alternatively, to measure greenwashing more directly, Hail et al. (2021) build on Sautner et al. (2023a) by examining a manager's 'excessive' discussion of climate change issues in response to analyst questions.

3.2. Alternative Approaches

Pattern-based sequence classification is a powerful tool for developing text-based measures. It is not a fail-safe method, though. Generating a vocabulary from 'authoritative' texts (i.e., the training library) has some pitfalls, some of which were mentioned above. Pattern-based sequence classification is likely to generate poor results when there is a discrepancy between the vocabulary used in the authoritative text and the documents used for scoring. For example, Sautner et al. (2023a) discuss how the language used to discuss climate change in policy documents from institutions such as the IPCC (Intergovernmental Panel on Climate Change) or the UN differs from the way firm managers talk about the topic in earnings calls. In addition, when interested in targeted aspects of an issue (e.g., regulatory vs. physical climate risk), putting together appropriate topic-specific libraries can be challenging (see Hassan et al., 2019). The scope and implications change (rapidly) for some unfolding events (e.g., COVID-19 or economic sanctions on Russia). Appropriate training materials are limited when researchers wish to study such shocks. This circumstance also poses a challenge to pattern-based sequence classification. We present two alternative approaches that have shown to be helpful in such cases.

3.2.1. Computer-assisted keyword discovery

Originating in the Political Sciences, keyword discovery provides a resourceful addition to approach unstructured text (King et al., 2017).[15] The advantage of this approach lies in recognizing that while humans have difficulties producing comprehensive word lists associated with a given topic, they can reliably know *some* of the vocabulary associated with a topic. These seed words need to be unambiguous keywords (e.g., for climate change: 'global warming,' 'carbon emission,' or 'Kyoto protocol') that experts could quickly agree to represent the topic. Keyword discovery expands the list by identifying the remainder of words in the vocabulary of interest by their co-occurrence with seeds. This process can be reiterated to learn higher-order co-occurrences.

Keyword discovery helps in cases where the available authoritative text (considered for the training library) employs a different language (more technical or more formal) than is used in the documents of interest. After agreeing on the seed words, the new keywords are 'discovered' from those documents. Thus, if the interest is to find keywords for disruptive technologies in earnings call transcripts, the seed words identify sentences in the transcript that are then considered for determining whether they contain new technology keywords.

[15] Unlike topic modeling such as LDA or CorEx, keyword discovery is not an algorithmic method but a computer-assisted method. This means that it does not rely on a fixed set of rules to generate keywords and topics but rather on a combination of statistical techniques and human input. The human input can come in the form of providing feedback, refinement, or domain knowledge to the method. The statistical techniques can include training classifiers, extracting information from their mistakes, and summarizing results with Boolean search strings. The goal is to suggest keywords that are relevant and informative for the user's task rather than to discover the latent structure or meaning of the text corpus.

Extracting relevant information from textual sources is challenging when the topic of interest is (technically) complex, fast-moving, and ambiguous. Resources for training machine learning algorithms tend to be inherently scarce then, limiting their effectiveness (Webersinke et al., 2021). Keyword discovery can overcome these difficulties as in Sautner et al. (2023a), who use the algorithm to identify new keywords based on climate change seeds. They report that seeds such as wind farms identify locations and facility names for renewable energy in earnings call transcripts that would have been virtually impossible to find otherwise.[16]

3.2.2. Word embedding

Certain topics, such as corporate culture, do not have specific word combinations but instead use phrases that are commonly encountered in novels, news, or professional publications. This makes training libraries and keyword discovery inapplicable. However, an algorithm that learns the meaning of words from their context can help identify different word combinations used in distinct contexts but with the same purpose. For example, when an algorithm recognizes that 'shoulder-to-shoulder' or 'hand-in-glove' are synonymous with teamwork, pertinent word combinations for cultural values can be developed.

The word embedding model is based on the insight that words that co-occur with the same neighboring words are likely to have similar meanings (Harris, 1954). This model identifies synonyms from common adjacent words and learns the meaning of a specific word via a neural network that 'reads' through textual documents and predicts its neighboring words.

Word embedding is a semi-supervised model similar to keyword discovery. Researchers must provide a small set of keywords related to the concepts of interest. For example, Li et al. (2021) used initial keywords, or seed words, developed by Guiso et al. (2015) by reading corporate values statements on firm web pages. Similarly, Bloom et al. (2021) collected keywords for disruptive technologies from patent award documents and then applied them to earnings call transcripts. To expand the disruptive technologies word list, Bloom et al. used word embedding to find terms with similar meanings to their initial disruptive technology seeds.

Pretrained language models, which have learned the representation of words and patterns of common language from large amounts of textual data, have become prominent in recent years. However, these gains in accuracy are not necessarily obtained when processing text of niche languages that use specific wording or when the wording can vary by source (Webersinke et al., 2021). For example, technical language is pervasive in climate change discussions, and the language used can vary among policymakers, journalists, and financial market participants. Ongoing work pretrains NLP language models on domain-specific text, such as FinBert, a pretrained language model for financial sentiment analysis (Araci, 2019; Huang et al., 2022).

A common feature among the algorithms discussed thus far is that researchers must provide initial inputs to guide the algorithms, such as the choice of training libraries or a small set of initial keywords. This limited human intervention is a great advantage of these methods, but researchers should report perturbation tests of their initial choices to convince readers that this intervention is not driving outcomes. In the next section, we offer more strategies to convince readers of the validity of text-based method outputs.

4. Establishing Empirical Credibility and Impact for Text-Based Measures

Empirical credibility is crucial when introducing a new text-based measure in accounting and finance research. Researchers must validate the measure and show that it captures economically

[16] As a result, the expanded keyword list may contain terms seemingly unrelated to the topic of interest.

significant variation at the appropriate level of analysis, typically at the firm level. This section delves into these aspects.

Newly proposed constructs often raise concerns about measurement error. New measures should noticeably improve hypothesis testing compared to existing data. Researchers must present a convincing argument that such improvements are achievable. In computational linguistics, accuracy statistics and performance metrics are emphasized, which are important but less critical in accounting and finance. Ultimately, the value of a new measure lies in providing additional economic insights.

4.1. Validity

This section explores various strategies for validating text-based measures. Validity refers to how accurately a text-based metric measures what it claims to (Nunnally & Bernstein, 1994). We examine face validity, convergent and discriminant validity, and human audit studies. These internal validity measures assess the quality of an investigation. Generalizability or external validity is often less of a concern as text-based studies in accounting and finance usually involve large samples (of company disclosures).

Face validity. Establishing face validity is a fundamental step in validating new measures, as demonstrated in recent work (Andreicovici et al., 2020; Bloom et al., 2021; Hassan et al., 2019, 2023a, 2023b; Sautner et al., 2023a). Researchers must show that their algorithms pass basic checks and build their measure from the ground up. For example, when assessing a firm's exposure to human rights abuse concerns, they must first identify relevant word combinations in firm disclosures. Next, they should examine text fragments containing these word combinations, ensuring meaningful discussions of human rights issues. Presenting both top and bottom-ranked scores helps avoid cherry-picking.

At the entity level, face validity can be established by documenting plausible patterns of the aggregated metric over time or demonstrating the distribution of average scores across industries makes intuitive sense. For example, the mean of a text climate change exposure measure should probably be high in 'brown' and low in environmentally friendly industries. A variance decomposition analysis can reveal whether there is sufficient firm-level heterogeneity in the new metric after accounting for time and sector fixed effects.

Human audit study. A human audit involves trained 'judges' evaluating the output of a machine-learning algorithm. In the gold standard human audit described by Baker et al. (2016), auditors reviewed over 12,000 newspaper articles using a 65-page guide to measure economic policy uncertainty in the U.S. The authors approach this task by counting the frequency of a small set of specific keywords related to economic policy in articles published in major U.S. newspapers across an extended period (beginning in the 1900s). About 20 percent of all articles are reviewed by multiple auditors. This resource-intensive process took over 18 months but significantly enhanced the credibility of their Economic Policy Uncertainty measure.[17] While demands for such verification may diminish as machine learning becomes more accurate and familiar, researchers must weigh the resource investment needed for conducting an audit study against the precision of results obtained. Often, limited-scale, iterative auditing processes improve algorithm accuracy and provide sufficient validation without committing to a full-scale assessment.[18]

---

[17]Details about the approach are given in an online supplement accessible on Professor Davis's homepage: https://stevenjdavis.com/research.

[18]Documentation is critical to support empirical credibility. The guiding principle is to carefully document the steps involving researchers' discretion. For machine-learning approaches in which researchers have limited discretion, documentation is more about version control of coding (e.g., pattern-based sequence classification). However, the documentation requirements will be more onerous for human audits. Researchers need to quantify the magnitude of

Convergent and discriminant validity. Convergent validity refers to the correlation between two theoretically related measures (Campbell & Fiske, 1959), while discriminant validity tests demonstrate the differentiation between distinct measures. Convergent validity tests allow researchers to demonstrate the connection between their new measure and existing measures of the same construct. For example, if a researcher measures 'human rights violations' using earnings call transcripts, they can also compute a similar score using the firm's annual report; the two measures should be positively correlated.

In the spirit of convergent validity, Jamilov et al. (2021) use the Privacy Rights Clearinghouse (PRC) database to identify reported cyber incidents and match these with their earnings call cybersecurity risk metric. These related but distinct measures should both be connected to the underlying construct of a firm's true cybersecurity risk. Utilizing data from a different database, like the PRC, adds credibility to the test. Hassan et al. (2019) correlate their firm-level 'political risk' metric with an index of regulatory constraints (Al-Ubaydli & McLaughlin, 2017), which is a sector-year score based on the relevance of a section of the Code of Federal Regulations (CFR) to a specific industry, multiplied by the number of occurrences of restrictive words in the CFR. Although these two scores differ and originate from separate text sources, their positive association increases the plausibility that they both relate to the underlying construct of true political risk.

Illustrative case studies: the proof of the pudding. Anton Chekhov's principle of *show, don't tell* applies equally to accounting and finance research. To demonstrate the effectiveness of a new text-based measure, one can examine a single firm's score over time, correlating peaks and troughs with significant events, potentially sourced from the original document. For instance, Hassan et al. (2019) report a power company case study, which illustrates various state-level, national, and legal actions corresponding to significant points in the firm's measured political risk time series. Jamilov et al. (2021) employ the 2017 Equifax and 2014 Target data breaches to demonstrate spikes in their cyber security metric.

4.2. *Strategic Incentives to Disclose: Threat Or Opportunity?*

Accounting researchers may question the validity of using disclosures to develop text-based measures for key issues facing firms, given numerous studies illustrating strategic reporting motives and communication obfuscation between executives and stakeholders (Armstrong et al., 2010; Dechow et al., 2010; Fields et al., 2001; Healy & Palepu, 2001).[19] Despite this evidence, these disclosures can be employed to study firms. In the first place, perhaps, textual analysis tools offer new opportunities to document *how* management acts strategically in its disclosures. Recent work has examined for this reason, word usage, tone, sentiment, readability, obfuscation, withholding strategies, and other communication aspects that suggest strategic objectives (e.g., Bushee et al., 2018; Gow et al., 2021; Hail et al., 2021).[20]

Beyond the study of strategic disclosure per se, when using these documents to extract information about economically significant firm events, emphasizing management's incentives to misrepresent disclosures will not advance the field. Instead, adjusting for these incentives in empirical design is feasible. Just as separating innate components from discretionary parts in reported earnings is a standard approach in accounting, controlling for misrepresentation incentives, as done by Hassan et al. (2019) when including prior period financial performance controls

---

measurement errors and present false negatives and positives separately. Documenting training procedures and materials for auditor training is also preferable.

[19]However, Ball (2013) offers relevant observations regarding the prevalence of these practices.

[20]See also Brown et al. (2020); Hobson et al. (2012); Larcker and Zakolyukina (2012); Mayew and Venkatachalam (2012).

in their regressions, can be effective.[21] In addition to controlling for past performance to capture strategic incentives, computing general sentiment measures from disclosures can also control for overly positive portrayals by managers. An open question is whether these countermeasures sufficiently reduce the likelihood of managerial deception to justify data usage. Our preliminary answer is: yes, they do!

4.3. Economic Applications: The Value of Text-based Measures

Using text methods and developing metrics should lead to novel insights. Ideally, the starting point should be to address real-world problems where data limitations have hindered progress. Real effects may arise in (1) capital markets, (2) investments and operations, and (3) labor markets and hiring. We provide examples for each below.

In accounting and finance, researchers often begin by examining capital market pricing for new text-based measures. For instance, Jamilov et al. (2021) report asset pricing effects of their cyber security risk measure, with a single additional mention of cyber security risk decreasing weekly returns by approximately 4.3 basis points. They also demonstrate that cyber security incidents have negative ripple effects on peer firms' returns. Sautner et al. (2023b) study the pricing of climate change exposure, using forward-looking option-pricing methods to show that text-based measures of climate change exposure are priced, but the premium declines over time as investors associate exposure with higher opportunities and lower crash risk.

Beyond capital market effects, Andreicovici et al. (2020) investigate a complex issue in organizational economics, examining how measurement frictions (i.e., challenges in converting economic transactions into financial statement numbers) impact business practices. They find that firms with high measurement intensity experience lower total factor productivity, hiring, and investments. These negative consequences partially stem from firms with higher measurement intensity struggling to establish effective compensation contracts with top executives. Textual methods in this study provide insights into the extent of measurement frictions within firms.

Bloom et al. (2021) explore labor market effects of disruptive technology diffusion (measured using textual methods) and the development and influence of technology hubs on cities. High-skilled jobs tend to diffuse more slowly, enabling cities with related technologies to benefit from their invention over extended periods. Research universities are crucial in fostering technological breakthroughs and establishing hubs where high-skilled workers concentrate and innovate.

5. Expanding the Scope of Text-based Research: Potential Directions

We recognize the challenge of recommending new research directions, but with this disclaimer, we highlight some emerging trends.

5.1. Micro-foundations of Macro Questions

Macroeconomic models rely on assumptions about firms' and consumers' preferences, expectations, and actions. Testing these assumptions empirically can uncover critical shortcomings in model predictions and drive refinements. Data availability has been challenging, but textual analysis of company disclosures offers a promising path forward. For example, management forecasts have been widely studied in accounting literature (see Beyer et al. (2010) for an overview). Although the literature primarily focuses on numerical sales growth and earnings

[21] Note that these methods to separate innate from discretionary components are subject to econometric concerns (see, e.g., Chen et al., 2018).

expectations, contextual information can be obtained from thoroughly parsing company textual documents.

Accounting research has highlighted how firm-level shocks can interact with aggregate shocks to suppress economy-wide hiring (Kalay et al., 2018). Gaining a deeper understanding of the underlying mechanisms involves examining various shocks firms face. Firm-level demand and supply-side shocks may amplify aggregate uncertainty differently and with varying strengths. Text data can help researchers understand firm-level shock heterogeneity and management responses (Hassan et al., 2023a).

Studying executives' expectations and their economic impacts is a flourishing area in literature (Bordalo et al., 2020; Gennaioli et al., 2016). The wealth of data available from company disclosures about managerial thought processes will likely yield many new insights. In particular, impromptu disclosures (such as earnings calls) can reveal subtle aspects. Analysts often request management to 'add some color' to their predictions, explicitly seeking soft information.

Flynn and Sastry (2022) offer a reverse perspective (from macroeconomy to microfoundations), showing that firms exhibit attention cycles to the macroeconomy, with peak attention during economic downturns due to the higher stakes in making accurate decisions. Attention to the macroeconomy is measured using textual data from conference calls.

GDP predictions are essential for capacity and stocking planning. Managers in better-managed firms seem to make more accurate future GDP predictions (Bloom et al., 2021). Next, we discuss how text data can be used to learn about a firm's internal operations and provide new broad-sample measures to identify such well-managed firms.

5.2. Using Text to Learn More About the Inner Firm

Management accounting researchers, lacking a COMPUSTAT equivalent, have been resourceful, investing significant effort in collecting survey data, obtaining private access to company records, or designing field experiments. Text data offers opportunities to complement these approaches. For instance, insight into a company's hierarchy and labor practices can be gleaned by examining job opening texts for clues. Li et al. (2014) demonstrate how communication patterns within a firm can be studied by observing executive interactions during conference calls. More broadly, economic models suggest a central role for language in understanding firms and their boundaries (Crémer et al., 2007). These models provide theoretical motivation for investigating the complexity of language used in firm communication (Dzieliński et al., 2021; Grennan, 2020). Other work highlights the role of 'clarity' in communicating relational contract terms (Gibbons & Henderson, 2012). The question here is how managers communicate promises about future behavior so that counterparties, such as employees, suppliers, or customers, understand their intentions. Conference calls are an excellent venue to learn how managers do this and whether they do so effectively. Much of the interaction in these calls is devoted to making claims about the future and doing so credibly.

A contemporary setting where relational contracts are likely important is in working-from-home arrangements. As employees and their superiors adapt to the lack of direct supervision, incentives, performance measures, and the firm's culture must adjust. Text from job ads and employment contracts can illuminate how this transformation is achieved. Contracts are a text source that should not be overlooked. They can provide details about dealings with suppliers and major customers and the compensation packages of senior management. Earlier work has recognized these possibilities (see, e.g., Costello, 2013), but from a textual perspective, much remains to be explored. Similarly, despite the investigation of debt contracts and covenants[22] , textual

[22] See Armstrong et al. (2010) for a review.

analysis of these documents will likely reveal further insights, such as the role of accounting information (Andreicovici et al., 2020).

5.3. Text Analysis for 'Traditional' Accounting Research Topics

Perhaps the most direct way text analysis can interest accounting researchers is by investigating 'traditional' disclosure questions. Current computational capabilities promise a systematic assessment of a firm's communication process, its dissemination to relevant stakeholders, stakeholders' responses, and management's response to those reactions. This allows for exploring interactions between financial and non-financial communication and how technology has disrupted prior engagement methods with external parties. Corporations and executives tweet (Crowley et al., 2021), and their social media presence offers more granular data on disclosure choices than ever. Ideas about 'management learning from market reactions' (Edmans et al., 2015) can be subjected to more rigorous tests. Theories about this type of learning may need to accommodate stock prices reflecting retail investor responses to a CEO's Facebook musings about their hobbies. Stakeholder responses can be harvested from analyst reports, media commentary, social media activity, or conference calls. Textual analysis allows examination of the full range of corporate communications and responses.

But possibilities extend beyond disclosure studies. Influential work on earnings properties, such as conservatism and value relevance, has employed relatively simple regression models correlating a few capital market and financial statement variables. However, if credible, conservative accrual choices should align with conservative disclosure choices. Thus, when ranking firms on their Basu-type conservatism, this ranking should correspond to a similar propensity to disclose bad news.

Other approaches use financial data to learn which textual disclosures impact capital markets. For example, Gad et al. (2022) employ lasso regressions to understand which words in conference calls of financial institutions and borrowers are priced in debt markets. This approach illustrates that economists can leverage tools developed in computational linguistics. The focus on text can be complemented by drawing inferences from market and pricing responses.

5.4. Using Text for Identification

Nonrandomization complicates causal inferences in observational studies. Researchers in accounting and finance have successfully exploited settings with credibly exogenous changes to the variable of interest. These settings then allow for applying research designs (e.g., difference-in-difference or regression discontinuity) that demonstrate causal effects. The lack of exogenous shocks often limits these approaches. A common alternative strategy is matching, i.e., finding untreated firms similar to treated ones. However, matching on 'observable' covariates is problematic, as important covariates are often unobservable or difficult to measure. Leveraging text data in the matching procedure can improve matching quality by incorporating *latent* confounders (Roberts et al., 2020; Veitch et al., 2020).[23] In the social science context, consider a researcher who wants to estimate the effect of ethnicity on college admissions. By matching applicants' profiles and essays, she can improve comparability and more accurately identify the impact of ethnicity. While existing matching approaches based on structured data (e.g., propensity score matching) are ill-equipped to handle high-dimensional text data, this problem can be addressed by incorporating dimensionality reduction in the matching procedure (Mozer et al., 2020; Roberts

[23] When confounders are well-understood but empirical proxies are missing, supervised/semi-supervised methods discussed in Section 3 (i.e., pattern-based sequence classification, computer-assisted word discovery, and word embedding) offer relief.

et al., 2020; Veitch et al., 2020). However, one concern is that the output of dimensionality reduction algorithms (e.g., topic modeling) is often difficult to interpret and offers little insight into potential confounders.

Instead of relying on a black-box model, Zeng et al. (2022) develop an alternative approach to uncover interpretable confounders. They study the effect of cancer treatments on survival outcomes. In their setting, clinical notes provide text data that can be used to identify a patient's underlying health condition, which determines the type of cancer treatment and treatment outcome. They first use simple NLP techniques to identify frequently occurring terms (applying tf*idf). Next, they use a prediction model (i.e., the Lasso model) to select terms predictive of both treatment assignment and treatment outcomes. For example, if the word 'bladder' has significant effects in both prediction models, this potential confounder must be controlled to make a credible causal inference. This approach is particularly valuable when uncovering potential confounders provides additional insights and helps establish the credibility of the strategy to identify the causal effect. To date, no empirical applications of using text for identification in finance and accounting have been identified, but exciting opportunities exist for text-based identification.

5.5. The Emergence of Large Language Models: Generative AI in Accounting

Earlier, we highlighted the advantages of using simple text analysis methods, arguing that they are transparent and easy to understand and interpret. For example, basic conditional word counts can measure how important a topic is within a text based on its frequency in the conversation. However, the recent development of advanced language models like Generative Pre-trained Transformers (e.g., ChatGPT and Bing.AI), introduced in November 2022, may shift the preference towards more powerful but less transparent algorithms. These models are excellent at sentiment analysis, summarization, and complex language comprehension because they can capture the relationships among words, sentences, and paragraphs.

Applications are just starting to emerge. So far, ChatGPT has been used to predict stock returns by analyzing the sentiment of news headlines related to a company (Lopez-Lira & Tang, 2023). A notable application in accounting research is Kim et al. (2023), which explores ChatGPT's ability to help investors extract value-relevant information from potentially 'bloated' disclosures. The study uses AI to summarize Management Discussion and Analysis disclosures and earnings calls. The authors define *bloat* as the degree to which the AI's summary reduces the original disclosure text and show that bloat is associated with negative capital market outcomes, such as lower stock price efficiency. ChatGPT also shows skill in identifying themes within the text and generating concise summaries of related discussions. Kim et al. (2023) suggest a way to detect ESG-focused discussions in disclosures using specific prompts that avoid possible problems, such as the language model making up answers when the text does not provide enough information (e.g., because ESG is not discussed).

6. Conclusions

This article explores how accounting and finance researchers can utilize text data to contribute to broader debates in economics and related fields. Although numerous social issues offer opportunities for accounting researchers, the community has been relatively subdued in addressing these problems, often considering them beyond the scope. The accountants' expertise in measuring and verifying complex business transactions, as well as their profound understanding of information production and consumption, raises questions about why they are not actively engaging with broader societal issues. Employing text as data may allow accountants to make valuable contributions to these discussions.

Our work can be summarized as follows: We propose methods for accounting researchers to formulate research questions with broad appeal and impact and introduce underutilized text sources for exploration. We discuss techniques for processing raw unstructured text into data and advocate for simplicity in text-based research. We offer guidance on presenting text-based evidence to establish empirical credibility and recommend strategies for demonstrating that a newly developed metric addresses economically significant questions.

Measurement is often the central challenge when tackling societal problems, but text data from various sources can help address this issue. We concentrate on pattern-based sequence classification, which can classify unstructured data into diverse economic topics. We emphasize the approach's conceptual simplicity, broad applicability, and minimal noise from researcher interventions. Moreover, we present alternative methods for cases where appropriate training materials use technical language that differs significantly from everyday conversation or conference calls.

Empirical credibility is essential for impact. Simple methods enable readers to comprehend how new measures are developed. Furthermore, we underline the empirical strategies for establishing a measure's validity. We outline useful practices that address concerns about the quality of text data, including whether managers' selective disclosure incentives contaminate data based on voluntary disclosures.

In conclusion, we spotlight four promising ideas for broadening the scope of text-based research. Accounting researchers can employ text data to construct micro-foundations for addressing macroeconomic questions or vice versa. Text data can complement internal data sources for management accounting, providing valuable insights into a firm's internal operations. Reexamining 'traditional' accounting research topics through the lens of computational linguistics may yield fresh perspectives. Lastly, text-based identification can be enhanced by improving matching quality. We also discuss the tantalizing new developments in generative AI models, which dramatically increase the power of textual analysis.

Acknowledgements

We thank Beatriz Garcia Osma (Editor) for inviting our thoughts on this subject on the occasion of *European Accounting Review*'s 30th anniversary and for constructive feedback on an earlier draft. We thank an anonymous reviewer, Margaret Abernethy, Mahmoud Gad, Ahmed Tahoun, Ruishen Zhang, and participants of the EAR 30 Years EAA Online Seminars, Ph.D. course participants at Columbia, New York, and Deakin Universities for their valuable comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Hung and Van Lent gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft Project ID 403041268 – TRR 266. Van Lent is also grateful for the financial support from the Institute for New Economic Thinking (INET).

ORCID

Jihun Bae http://orcid.org/0000-0003-1580-8718

Chung Yu Hung http://orcid.org/0000-0003-0805-5260

Laurence van Lent http://orcid.org/0000-0002-9354-0932

References

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). *Ai and jobs: Evidence from online vacancies* (Tech. Rep.), National Bureau of Economic Research.

Al-Ubaydli, O., & McLaughlin, P. A. (2017). RegData: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012. *Regulation & Governance*, *11*, 109–123. https://doi.org/10.1111/rego.v11.1

Allee, K. D., & DeAngelis, M. D. (2015). The structure of voluntary disclosure narratives: Evidence from tone dispersion. *Journal of Accounting Research*, *53*(2), 241–274. https://doi.org/10.1111/joar.2015.53.issue-2

Andreicovici, I., van Lent, L., Nikolaev, V. V., & Zhang, R. (2020). Accounting measurement intensity. SSRN Scholarly Paper ID 3639267, Social Science Research Network, Rochester, NY.

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.

Armstrong, C. S., Guay, W. R., & Weber, J. P. (2010). The role of information and financial reporting in corporate governance and debt contracting. *Journal of Accounting and Economics*, *50*(2-3), 179–234. https://doi.org/10.1016/j.jacceco.2010.10.001

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Ball, R. (2013). Accounting informs investors and earnings management is rife: Two questionable beliefs. *Accounting Horizons*, *27*(4), 847–853. https://doi.org/10.2308/acch-10366

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2-3), 296–343. https://doi.org/10.1016/j.jacceco.2010.10.003

Bierman, H. (1963). Measurement and accounting. *The Accounting Review*, *38*(3), 501–507.

Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J., & Tahoun, A. (2021). *The diffusion of disruptive technologies*, Working Paper 28999, National Bureau of Economic Research. Series: Working Paper Series.

Bloom, N., Kawakubo, T., Meng, C., Mizen, P., Riley, R., Senga, T., & Van Reenen, J. (2021). *Do well managed firms make better forecasts?* (Tech. Rep.), National Bureau of Economic Research.

Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2023). Textual analysis in accounting: What's next? *Contemporary Accounting Research*. *40*(2), 765–805. https://doi.org/10.1111/1911-3846.12825

Bochkay, K., Chychyla, R., & Nanda, D. (2019). Dynamics of CEO disclosure style. *The Accounting Review*, *94*(4), 103–140. https://doi.org/10.2308/accr-52281

Bochkay, K., Hales, J., & Chava, S. (2020). Hyperbole or reality? Investor response to extreme language in earnings conference calls. *The Accounting Review*, *95*(2), 31–60. https://doi.org/10.2308/accr-52507

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, *110*(9), 2748–2782. https://doi.org/10.1257/aer.20181219

Bowman, E. H., & Haire, M. (1976). Social impact disclosure and corporate annual reports. *Accounting, Organizations and Society*, *1*(1), 11–21. https://doi.org/10.1016/0361-3682(76)90004-0

Brochet, F., Loumioti, M., & Serafeim, G. (2015). Speaking of the short-term: Disclosure horizon and managerial myopia. *Review of Accounting Studies*, *20*(3), 1122–1163. https://doi.org/10.1007/s11142-015-9329-8

Brochet, F., Miller, G. S., Naranjo, P., & Yu, G. (2019). Managers' cultural background and disclosure attributes. *The Accounting Review*, *94*(3), 57–86. https://doi.org/10.2308/accr-52290

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research*, *58*(1), 237–291. https://doi.org/10.1111/joar.v58.1

Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, *49*(2), 309–346. https://doi.org/10.1111/joar.2011.49.issue-2

Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research*, *56*(1), 85–121. https://doi.org/10.1111/1475-679X.12179

Bushee, B. J., & Huang, Y. J. (2021). Do analysts and investors efficiently respond to managerial linguistic complexity on conference calls? Available at SSRN 3405086.

Cai, W., Kaplan, R. S., & Zou, Y. (2021). *Preparing for a rainy day: Firm capital and novel risk*, SSRN Scholarly Paper ID 3987659, Social Science Research Network, Rochester, NY.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Campbell, D., Loumioti, M., & Wittenberg-Moerman, R. (2019). Making sense of soft information: Interpretation bias and loan quality. *Journal of Accounting and Economics*, *68*(2-3), 101240. https://doi.org/10.1016/j.jacceco.2019.101240

Chan, J. L. (1979). Corporate disclosure in occupational safety and health: Some empirical evidence. *Accounting, Organizations and Society*, *4*(4), 273–281. https://doi.org/10.1016/0361-3682(79)90018-7

Chen, W., Hribar, P., & Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research*, *56*(3), 751–796. https://doi.org/10.1111/1475-679X.2018.56.issue-3

Cohen, L., Malloy, C., & Nguyen, Q. (2020). Lazy prices. *The Journal of Finance*, *75*(3), 1371–1415. https://doi.org/10.1111/jofi.v75.3

Costello, A. M. (2013). Mitigating incentive conflicts in inter-firm relationships: Evidence from long-term supply contracts. *Journal of Accounting and Economics*, *56*(1), 19–39. https://doi.org/10.1016/j.jacceco.2013.02.001

Crémer, J., Garicano, L., & Prat, A. (2007). Language and the theory of the firm. *The Quarterly Journal of Economics*, *122*(1), 373–407. https://doi.org/10.1162/qjec.122.1.373

Crowley, R. M., Huang, W., & Lu, H. (2021). *Executive tweets*, SSRN Scholarly Paper ID 3975995, Social Science Research Network, Rochester, NY.

Darendeli, A., Law, K. K., & Shen, M. (2022). Green new hiring. *Review of Accounting Studies*, *27*(3), 986–1037. https://doi.org/10.1007/s11142-022-09696-y

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies*, *20*(2), 639–673. https://doi.org/10.1007/s11142-014-9309-4

Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, *50*(2-3), 344–401. https://doi.org/10.1016/j.jacceco.2010.09.001

Dzieliński, M., Wagner, A. F., & Zeckhauser, R. J. (2021). *CEO clarity*, RWP17-017, Swiss Finance Institute Research Paper, (No. 17–13).

Edmans, A., Goldstein, I., & Jiang, W. (2015). Feedback effects, asymmetric trading, and the limits to arbitrage. *American Economic Review*, *105*(12), 3766–3797. https://doi.org/10.1257/aer.20141271

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, *33*(3), 1184–1216. https://doi.org/10.1093/rfs/hhz072

Fields, T. D., Lys, T. Z., & Vincent, L. (2001). Empirical research on accounting choice. *Journal of Accounting & Economics*, *31*(1-3), 255–307. https://doi.org/10.1016/S0165-4101(01)00028-3

Flynn, J. P., & Sastry, K. (2022). Attention cycles. Available at SSRN 3592107.

Frankel, R., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, *37*(1), 133–150. https://doi.org/10.2307/2491400

Frazier, K. B., Ingram, R. W., & Tennyson, B. M. (1984). A methodology for the analysis of narrative accounting disclosures. *Journal of Accounting Research*, *22*, 318–331. https://doi.org/10.2307/2490713

Fritsch, F. (2022). *Listening to twitter: Helpful for managerial forecasting?* (Tech. Rep.), Frankfurt am Main.

Fritsch, F., Zhang, Q., & Zheng, X. (2022). *Responding to climate change crises: Firms' tradeoffs* (Tech. Rep.), Frankfurt am Main.

Gad, M., Nikolaev, V. V., Lent, L. Van, & Tahoun, A. (2022). *Firm-level political risk and credit markets*, Working Paper. TRR 266 Accounting for Transparency Working Paper Series No. 87.

Gennaioli, N., Ma, Y., & Shleifer, A. (2016). Expectations and investment. *NBER Macroeconomics Annual*, *30*(1), 379–431. https://doi.org/10.1086/685965

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gibbons, R., & Henderson, R. (2012). Relational contracts and organizational capabilities. *Organization Science*, *23*(5), 1350–1364. https://doi.org/10.1287/orsc.1110.0715

Gow, I. D., Larcker, D. F., & Zakolyukina, A. A. (2021). Non-Answers during conference calls. *Journal of Accounting Research*, *59*(4), 1349–1384. https://doi.org/10.1111/joar.v59.4

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting & Economics*, *40*(1-3), 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

Grennan, J. (2020). *Communicating culture consistently: Evidence from banks*, SSRN Scholarly Paper ID 3350645, Social Science Research Network, Rochester, NY.

Guiso, L., Sapienza, P., & Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, *117*(1), 60–76. https://doi.org/10.1016/j.jfineco.2014.05.010

Hail, L., Kim, S., & Zhang, R. X. (2021). How do managers greenwash? Evidence from earnings conference calls.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hassan, T. A., Hollander, S., Lent, L. van, & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, *134*(4), 2135–2202. https://doi.org/10.1093/qje/qjz021

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2023a). *Firm-level exposure to epidemic diseases: COVID-19, SARS, and H1N1*, Review of Financial Studies. Forthcoming. https://doi.org/10.1093/rfs/hhad044

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2023b). The global impact of brexit uncertainty. *Journal of Finance*, Forthcoming. http://dx.doi.org/10.2139/ssrn.3840164

Healy, P. M., & Palepu, K. (2001). Information asymmetry, corporate disclosure and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, *31*, 405–440. https://doi.org/10.1016/S0165-4101(01)00018-0

Hines, R. D. (1988). Financial accounting: In communicating reality, we construct reality. *Accounting, Organizations and Society*, *13*(3), 251–261. https://doi.org/10.1016/0361-3682(88)90003-7

Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, *50*(2), 349–392. https://doi.org/10.1111/joar.2012.50.issue-2

Hollander, S., Pronk, M., & Roelofsen, E. (2010). Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, *48*(3), 531–563. https://doi.org/10.1111/joar.2010.48.issue-3

Homburger, R. H. (1961). Measurement in accounting. *The Accounting Review*, *36*(1), 94–99.

Huang, K., Li, M., & Markov, S. (2020). What do employees know? Evidence from a social media platform. *The Accounting Review*, *95*(2), 199–226. https://doi.org/10.2308/accr-52519

Huang, A. H., Wang, H., & Yang, Y. (2022). Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Impink, S. M., Prat, A., & Sadun, R. (2021). *Communication within firms: Evidence from CEO turnovers*, Working Paper 29042, National Bureau of Economic Research. Series: Working Paper Series.

Ingram, R. W., & Frazier, K. B. (1980). Environmental performance and corporate disclosure. *Journal of Accounting Research*, *18*(2), 614–622. https://doi.org/10.2307/2490597

International Accounting Standards Board (2017). *Disclosure initiative–principles of disclosure* (Tech. Rep., No. DP/2017/1), IFRS Foundation.

Jame, R., Johnston, R., Markov, S., & Wolfe, M. C. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, *54*(4), 1077–1110. https://doi.org/10.1111/joar.2016.54.issue-4

Jamilov, R., Rey, H., & Tahoun, A. (2021). *The anatomy of cyber risk*, Working Paper 28906, National Bureau of Economic Research. Series: Working Paper Series.

Jones, M. J., & Shoemaker, P. A. (1994). Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature*, *13*, 142.

Kalay, A., Nallareddy, S., & Sadka, G. (2018). Uncertainty and sectoral shifts: The interaction between firm-level and aggregate-level shocks, and macroeconomic activity. *Management Science*, *64*(1), 198–214. https://doi.org/10.1287/mnsc.2016.2581

Kalmenovitz, J. (2019). *Regulatory intensity and firm-specific exposure*, SSRN Scholarly Paper ID 3451344, Social Science Research Network, Rochester, NY.

Kim, A. G., Muhn, M., & Nikolaev, V. V. (2023, April). *Bloated disclosures: Can chatgpt help investors process information?* Chicago Booth Research Paper 23-07, University of Chicago, Becker Friedman Institute for Economics Working Paper.

King, G., Lam, P., & Roberts, M. E. (2017). Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science*, *61*(4), 971–988. https://doi.org/10.1111/ajps.2017.61.issue-4

Kirk, M. P., & Markov, S. (2016). Come on over: Analyst/investor days as a disclosure medium. *The Accounting Review*, *91*(6), 1725–1750. https://doi.org/10.2308/accr-51418

Klaassen, J. (1992). De winst is absoluut relatief (on)belangrijk. *MAB*, *5*, 202–203. https://doi.org/10.5117/mab.66.16777

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*(2), 495–540. https://doi.org/10.1111/joar.2012.50.issue-2

Lee, J. (2016). Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *The Accounting Review*, *91*(1), 229–250. https://doi.org/10.2308/accr-51135

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265–3315. https://doi.org/10.1093/rfs/hhaa079

Li, F., Minnis, M., Nagar, V., & Rajan, M. (2014). Knowledge, compensation, and firm value: An empirical analysis of firm communication. *Journal of Accounting and Economics*, *58*(1), 96–116. https://doi.org/10.1016/j.jacceco.2014.06.003

Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., & Lu, X. (2021). Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, *2*(4), 246–252. https://doi.org/10.1016/j.jnlssr.2021.10.003

Lopez-Lira, A., & Tang, Y. (2023, April). *Can chatgpt forecast stock price movements? Return predictability and large language models*, SSRN Working Paper id4412788, SSRN.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(14), 1187–1230. https://doi.org/10.1111/joar.2016.54.issue-4

Mahlendorf, M., Martin, M., & Smith, D. (2022). Innovative data–use-cases in management accounting research and practice. *European Accounting Review*. Forthcoming. http://dx.doi.org/10.2139/ssrn.4297954

Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, *86*(4), 1383–1414. https://doi.org/10.2308/accr-10034

Mayew, W. J. (2008). Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research*, *46*(3), 627–659. https://doi.org/10.1111/j.1475-679X.2008.00285.x

Mayew, W. J., Parsons, C. A., & Venkatachalam, M. (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior*, *34*(4), 243–248. https://doi.org/10.1016/j.evolhumbehav.2013.03.001

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, *67*(1), 1–43. https://doi.org/10.1111/j.1540-6261.2011.01705.x

Mozer, R., Miratrix, L., Kaufman, A. R., & Anastasopoulos, L. J. (2020). Matching with text data: An experimental evaluation of methods for matching documents and of measuring match quality. *Political Analysis*, *28*(4), 445–468. https://doi.org/10.1017/pan.2020.1

Neimark, M. (1983). How to use content analysis in historical research. *Accounting Historians Notebook*, *6*(2), 1.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*, McGraw-Hill Series in Psychology. McGraw-Hill.

Poshalian, S., & Crissy, W. J. (1952). Corporate annual reports are difficult, dull reading, human interest value low, survey shows. *Journal of Accountancy (pre-1986)*, *94*(2), 215.

Ranta, M., Ylinen, M., & Jävenpää, M. (2022). Machine learning in management accounting research: literature review and pathways for the future. *European Accounting Review*. Forthcoming. https://doi.org/10.1080/09638180.2022.2137221

Roberts, M. E., Stewart, B. M., & Nielsen, R. A. (2020). Adjusting for confounding with text matching. *American Journal of Political Science*, *64*(4), 887–903. https://doi.org/10.1111/ajps.v64.4

Sautner, Z., van Lent, L., Vilkov, G., & Zhang, R. (2023a). Firm-level climate change exposure. *The Journal of Finance*, *78*(3), 1449–1498. https://doi.org/10.1111/jofi.v78.3

Sautner, Z., van Lent, L., Vilkov, G., & Zhang, R. (2023b). Pricing climate change exposure. *Management Science*. Forthcoming. https://doi.org/10.1287/mnsc.2023.4686

Soper, F. J., & Dolphin, R. (1964). Readability and corporate annual reports. *The Accounting Review*, *39*(2), 358.

Suijs, J. (2007). Voluntary disclosure of information when firms are uncertain of investor response. *Journal of Accounting and Economics*, *43*(2-3), 391–410. https://doi.org/10.1016/j.jacceco.2006.10.002

van Zaanen, M., & Kanters, P. (2015). Automatic mood classification using tf*idf based on lyrics. In J. Downie and R. Veltkamp (Eds.), *11th International society for music information retrieval conference (ISMIR 2010)* (pp. 75–80). TiCC.

Veitch, V., Sridhar, D., & Blei, D. (2020). Adapting text embeddings for causal inference. In *Conference on uncertainty in artificial intelligence* (pp. 919–928). PMLR.

Watts, R. L. (1977). Corporate financial statements, a product of the market and political processes. *Australian Journal of Management*, *2*(1), 53–75. https://doi.org/10.1177/031289627700200104

Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2021). ClimateBert: A pretrained language model for climate-related text. arXiv:2110.12010.

Wu, X. (2020). *SEC regulations and firms*, SSRN Scholarly Paper ID 3625115, Social Science Research Network, Rochester, NY.

Zeng, J., Gensheimer, M. F., Rubin, D. L., Athey, S., & Shachter, R. D. (2022). Uncovering interpretable potential confounders in electronic medical records. *Nature Communications*, *13*(1), 1014. https://doi.org/10.1038/s41467-022-28546-8