Are Uncertain Firms Riskier?*

Fahiz Baba-Yara† Carter Davis‡ Fotis Grigoris§ Preetesh Kantak¶

June 15, 2023

Abstract

We use novel data covering 2 billion daily employee-article interactions across approximately 2 million firms to characterize firms' exposures to uncertainty in almost real-time. We find that, in the cross-section, firms that more intensely read about *financial* versus other uncertainty-related topics are those most exposed to changes in aggregate measures of economic uncertainty. Consistent with exposure to uncertainty being priced, public firms that spend more time reading these topics have a 2% higher cost of capital, translating into relatively low investments rates. Higher attention to financial uncertainty relates to 7% lower investment and 5% lower hiring on an annual basis.

*We are grateful to Greg Brown, Riccardo Colacito, Daniel Carvalho, Max Croce, Zhi Da, Jesse Davis, Andrew Ellul, Joseph Engelberg, Andrei Goncalves, Christian Heyerdahl-Larsen, Isaac Hacamo, Harrison Hong, Christian Lundblad, Stavros Panageas, Anna Pavlova, Jeffrey Pontiff, Jacob Sagi, Gill Segal, Noah Stoffman, Andrea Tamoni, Charles Trzcinka, Stijn Van Nieuwerburgh, Annette Vissing-Jorgensen, Zhenyu Wang, Ingrid Werner, and Guofu Zhou (discussant) for their valuable input. We would also like to thank conference and seminar participants at BI Norwegian, the 2023 Midwest Finance Association Conference, Indiana University—Bloomington, and the University of North Carolina—Chapel Hill. The authors further acknowledge the Indiana University Pervasive Technology Institute for providing HPC database, storage, and consulting resources that have contributed to the research results reported within this paper (Stewart, Welch, Plale, Fox, Pierce, and Sterling (2017)). Finally, we are indebted to our anonymous data provider for its generous support. All errors are our own.

†Indiana University Kelley School of Business. fababa@iu.edu
‡Indiana University Kelley School of Business. ckd1@iu.edu
§Indiana University Kelley School of Business. fgrigori@iu.edu
¶Indiana University Kelley School of Business. pkantak@iu.edu

Firms, households, and policymakers face the daunting task of deciding how to best allocate limited resources in an economic environment that evolves continuously and changes rapidly. To understand the decision-making underlying this task, a growing literature has turned towards analyzing the uncertainty associated with future economic conditions through the supply of news. Measuring uncertainty through news is particularly powerful in that it provides a real-time indicator of the salient economic events and narratives that economic agents are most concerned about at a given point in time (see, e.g., Baker, Bloom, and Davis (2016)). But while the *production* of uncertainty-related news is useful for understanding the concerns of the average reader, it does little to shed light on who is reading what types of news and why. In contrast, the *consumption* of uncertainty-related news would allow researchers to answer specific questions about the cross-section of agents: (1) Do individuals at organizations pay attention to uncertainty-related news?; (2) If so, does the type of news they read inform us about their exposure to uncertainty?; and (3) What plausible actions do they take to mitigate their exposure?

In this paper, we answer these questions for corporations (firms) by exploiting a granular dataset quantifying firm-employee demand for online content across more than two million firms, four thousand online publishers, and two billion daily user interactions. Using our dataset, we construct a novel, firm-level measure of uncertainty-related news consumption. When averaged across all firms, the measure is highly, but not perfectly, correlated ($\rho \approx 0.45$-0.70) with traditional measures of macroeconomic uncertainty, such as the EPU index itself, VIX, and the uncertainty measures of Ludvigson, Ma, and Ng (2021). As published news articles are an equilibrium outcome of consumers' preferences and the media's production technology (Mullainathan and Shleifer, 2005), it is reassuring but not altogether surprising to find that aggregate consumption of uncertainty-related news is correlated with other measures of uncertainty in the time series. We thus devote the bulk of our analysis to better understanding how cross-sectional differences in the amount and type of uncertainty-related news consumed relate to firm characteristics and firm economic outcomes.

Our key empirical result is to show that cross-sectional differences in the degree to which a firm is reading about financial uncertainty versus other business-related topics (henceforth, the "relative attention" of the firm) is highly informative about the firm's risk, mitigation activities, and future real outcomes. Notably, we show that firms that read more finance uncertainty-related news not only have higher implied costs of equity capital, but also, subsequently, partake in greater amounts of corporate hedging activity, invest less in capital, hire fewer or fire more workers, and have lower sales growth rates. The positive (negative) relationship between a firm's relative attention and its risk (real outcomes) is not explained by fixed differences across industries (e.g., the possibility the employees of riskier durable goods manufacturers read more about uncertainty than the employees of safer nondurable goods manufacturers), and exists *within* industry groups after we control for other important firm-level characteristics, such as a size, value, investment rates, and profitability. Taken together, these results highlight how a firm's consumption of uncertainty-related news provides a direct and timely indicator of each firm's risk and real policies.

The data underlying our relative attention measure is based on the employee reading of articles from a consortium of more than four thousand online content publishers (hereafter referred to as "the Consortium"). Each of these four thousand publishers—which span a variety of news publications (e.g., the Wall Street Journal, Forbes, and Bloomberg) and trade periodicals (e.g, Hart Energy and Step Stone)—provide the Consortium with comprehensive data on each user-article interaction (e.g., the URL of the article read, time of reading, external IP address). This allows the Consortium to link each of the over 2 *billion* user interactions on each calendar day between 2016 and 2022 to specific firms and employees. For instance, the Consortium can observe that ten unique users at Company A were reading the same article on the Wall Street Journal on a given day, while one user at Company B was reading two different articles on the same day.

The Consortium then deploys a state-of-the-art machine learning algorithm on the content of each article to decompose the article into its essential topics.[1] Returning to the previous example, the Consortium may determine that the Wall Street Journal article read by Company A was 30% related to "inflation" and 70% related to "FOMC". In contrast, the first article read by Company B was 100% related to "CPUs," while the second article was a 50:50 split between "CPUs" and "Cloud Computing." After decomposing each article into its topics, the Consortium aggregates topic interactions within a given firm and day to produce a daily dataset of firm-topic interactions. In our example, the Consortium would uncover that Company A (Company B) was predominately focused on the "FOMC" ("CPU") topic for that day.

[1]The Consortium's primary purpose is to generate a signal of user *intent* to purchase an underlying product or service. The objective of this machine learning algorithm is to generate a topic-specific signal so that a client of the Consortium can better direct sales, marketing and advertising dollars towards a firm whose topic-related intent is high. As such, the Consortium is economically motivated to (i) ensure its article decomposition is accurate and (ii) decompose articles into a diverse set of nearly 7,000 topics of interest to users.

While the set of news sources, articles, and topics covered in the Consortium's dataset is vast, we are primarily interested in the degree to which firms are paying attention to uncertainty-related news. In order to define a constrained set of uncertainty-related topics, we start with the articles underlying the EPU index of Baker et al. (2016). The publications represented are, among others, the Wall Street Journal, USA Today, and New York Times. We have the Consortium deploy its machine learning algorithm on these articles, which decomposes each article into the set of available topics. This identifies the set of topics underlying the EPU corpus of news on a given day. Across the sample from 2016 through 2022, approximately 600 topics qualify as being related to uncertainty. We then aggregate the reading of this subset of topics within a firm on a given day to obtain a firm-by-day measure of how intensely their employees are reading uncertainty-related news. Unlike the standard EPU index that is based on the number of articles featuring prominent key words, our measure of attention to uncertainty is related to a firm-employees' *consumption* of EPU-related news on a given day.

As noted, the set of news articles published are an equilibrium outcome of the media's production decisions and readers' preferences; as such the resultant panel of firm attention still confounds supply and demand for news. To distinguish between the two, we compute a topic-specific score using a procedure similar to that used in estimating *tf-idf* scores in computational linguistics (see, e.g., Gentzkow, Kelly, and Taddy, 2019, for an application in finance). Simply put, terms with high (low) *tf-idf* scores are those that are most (least) useful for differentiating the content of a given document from that of a cross-section of documents. Similarly, for a given week we compute a *topic frequency, inverse aggregate frequency* (henceforth, *tf-iaf*) score for each uncertainty-related topic. Topics that we assign as having a high (low) *tf-iaf* score in a given week are those that are the most (least) useful for differentiating what firms are paying attention to in the cross-section.[2] For instance, many firm-employees will read articles about inflation in the weeks when the Consumer Price Index (CPI) is reported by virtue of there being a higher supply of such articles during that time. Knowing that two firms — A and B — are reading about inflation in these weeks may tell us little about which firm is more exposed to uncertainty. Consequently, inflation-related topics will have low *tf-iaf* scores in these weeks. However, if Firm B also intensely reads other uncertainty-

[2]The primary difference between a *tf-idf* score and our *tf-iaf* score is that while the former measure is used to differentiate the content of *documents* in a corpus, our *tf-iaf* measure is used to differentiate firms based on their *attention* to various topics in the news.

related topics (e.g., interest rate swaps or duration management), while Firm A is not focused on these topics, then these topics will carry high *tf-iaf* scores. We then classify Firm B as plausibly more exposed to uncertainty than Firm A in those weeks. The same logic also underlies the use of *tf-idf* scores within the political risk measure of Hassan, Hollander, Van Lent, and Tahoun (2019).

Our measure of firm-level uncertainty reflects the proportion of a firm's total attention that is allocated to uncertainty-related topics, where all topics are weighted by their *tf-iaf* scores. With this *tf-iaf* weighted measure of uncertainty-news consumption in hand, we perform a simple validation exercise. If our *tf-iaf* adjusted measure of relative attention (henceforth, $ARA_{i,t}$) properly reflects a firm i's exposure to uncertainty, then one would assume that the uncertainty-related reading of these high $ARA_{i,t}$ would be more sensitive to fluctuations in uncertainty than the reading of low $ARA_{i,t}$ firms. That is, higher $ARA_{i,t}$ firms would have higher firm-employee reading βs to variation in uncertainty proxies. The results from this exercise are economically and statistically stark. When either VIX, EPU, or the Ludvigson et al. (2021) uncertainty measures increase, then the proportion of reading allocated towards uncertainty-related news increases by about 1.5 times more for high $ARA_{i,t}$ firms than low $ARA_{i,t}$ firms.

In short, the set of topics that have been upweighted (downweighted) to maximize cross-sectional differences in uncertainty-related reading seems to tell us quite a bit about the exposure of firms to uncertainty. A natural question is: how is the raw set of topics different from the *tf-iaf* weighted topics in terms of salience? Exploiting the granularity of topics available in our data, visual inspection shows that raw-reading topics have a mix of many types of uncertainty, such as political uncertainty (e.g. "economic inequality"), financial uncertainty (e.g. "currency futures"), and macroeconomic uncertainty (e.g. "inflation"). In contrast, the topics that best describe differences in the cross section focus mostly on financial uncertainty. These include topics such as corporate risk management (e.g., "duration management" and "liquidity and cash management"), hedging (e.g., "forex swaps" and "deposit insurance"), and regulatory risks (e.g., "Basel II" and "accounting standards board"). Our cross-sectional measure of uncertainty weights topics focused on financial topics so saliently that we refer to it as *financial* uncertainty reading.[3]

Since many of the important uncertainty-related topics firms are related to risk management, we

[3] In a placebo test, we also empirically verify that the proportion of raw, non-*tf-iaf* weighted uncertainty reading tells us little about a firm's β exposure to aggregate uncertainty.

next statistically test whether high $ARA_{i,t}$ firms are likely to partake, subsequently, in corporate hedging and compliance activity. First, we construct a proxy for hedging via firms' 10-Ks (see, e.g., Campello, Lin, Ma, and Zou, 2011, for more details on the measure) and show that the firms that read the most about uncertainty are 30% more likely to talk about hedging activity than the median firm in the sample. Second, we use the firm-level indices of regulatory burden from Kalmenovitz (2022) to show that firms that read the most about uncertainty have between a 2-10% higher regulatory burden than those that read the least. These findings are robust to a variety of controls and industry-by-date fixed effects. This finding is important in that many (e.g. Hassan et al., 2019) have linked hedging activity and regulatory burdens to industry exposure and specific macroeconomic events; this specification shows the relationship between reading and hedging holds even controlling for such events.

We then build on the fact that reading more uncertainty-related news tends to suggest a higher exposure to uncertainty by asking how the intensity of uncertainty-related news reading is associated with a firm's cost of capital. Firms that read the most about uncertainty in the cross-section have implied costs of capital that are roughly 2% higher than firms that read the least about uncertainty. This difference in the cost of capital across firms is also not driven by common shocks at a given point in time (e.g., the onset of the COVID-19 pandemic) or differences in the cost of capital across industries (i.e., the fact the durable goods manufacturers are riskier than non-durable goods manufacturers). Moreover, the association between the intensity of uncertainty-related reading and the cost of equity survives when we control for industry-by-time fixed effects, and a battery of firm-level characteristics typically associated with risk (e.g., size, leverage, and profitability).[4]

Finally, having shown that reading important uncertainty-related news is related to the cost of capital, we also show that attention to uncertainty predicts the future corporate decisions of firms. For instance, the firms with the highest relative attention measures invest about 3.50% (1.80%) per quarter less in physical capital (inventories) and have hiring growth rates that are 5% per annum lower than firms with the lowest attention to uncertainty. We establish these results by estimating firm-level panel regressions that not only control for a comprehensive set of observable firm characteristics that are known to predict real firm outcomes but also control for a variety of

[4]We find qualitatively similar results when we examine differences in the cost of capital across firms using a simple portfolio formation procedure and realized returns rather than a regression approach based on analyst-implied costs of capital.

fixed effects that account for unobservable differences between industries and across time. Since our most comprehensive specification controls for industry-by-time fixed effects, our results hold when comparing two firms operating in the same industry at a given point in time.

Related literature. This paper contributes to several strands of literature in economics and finance. Our focus on a firm's attention to uncertainty is motivated by the growing literature on the largely negative effects of macroeconomic, financial market, and policy uncertainty on firm-specific and economy-wide outcomes (see, inter alia, Bernanke, 1983; Bloom, 2009, 2014; Fernández-Villaverde, Guerrón-Quintana, Kuester, and Rubio-Ramírez, 2015; Jurado, Ludvigson, and Ng, 2015). This literature, however, often focuses on aggregate outcomes. The lack of clear proxies of exposure to uncertainty has hindered previous attempts to explore how uncertainty shocks have differential impacts across firms. This is because the preponderance of firm-level measures are either generated or inferred from (i) financial assets, which are priced by investors, not firms (e.g., Alfaro, Bloom, and Lin, 2018; Bali, Brown, and Tang, 2017), (ii) earnings call transcripts, which are an equilibrium outcome of investor and firm interactions (e.g., Hassan et al., 2019), and (iii) surveys, which are often stale subjective beliefs of a small subset of households, forecasters, or corporations (e.g., Altig, Barrero, Bloom, Davis, Meyer, and Parker, 2022). In contrast, our relative attention measure is available for each firm based on its own employee reading of relevant news in real-time. As it's based on reading of articles and specific topics it shares the rich level of granularity, but now in the cross-section of firms, commonly seen in other news-based measures (e.g., Ke, Kelly, and Xiu, 2019).

Recent research has turned to textual analysis of news articles to quantify the uncertainty associated with future economic conditions. For example, Baker et al. (2016) develop an index of Economic Policy Uncertainty (EPU) based on newspaper coverage frequency, which has become a widely-used measure of aggregate uncertainty. Other notable papers in this strand of literature include Ke et al. (2019), where the authors demonstrate the predictive power of news-based uncertainty measures for macroeconomic and financial outcomes. While these studies have advanced our understanding of the production of uncertainty-related news, they have not explored the consumption of such news at the firm level and its implications for firms' behavior and performance, which is the focus of our study.

Several studies have investigated firm-level exposure to uncertainty and risk. For example, Alfaro et al. (2018) focused on the role of financial assets in pricing uncertainty, while Hassan et al. (2019) analyzed earnings call transcripts to study firm-level uncertainty. Another strand of literature has used survey data to examine uncertainty (e.g., Altig et al. (2022)). While these studies have contributed valuable insights to the literature, they are limited by the nature of their data sources, which may not provide direct and timely indicators of each firm's exposure to uncertainty. Our study contributes to this literature by introducing a novel measure of firm-level news consumption, which offers a more direct and real-time indicator of firms' exposure to uncertainty and risk.

It is not a given that the employees of firms that are more exposed to financial uncertainty also read more about financial uncertainty topics. Indeed, Chinco, Hartzmark, and Sussman (2022) argue that people do not consider consumption risk in investment decisions, despite consumption-based asset pricing being a main workhorse model in asset pricing. Zingales (2015) emphasizes that academics tend to overstate the importance of finance. By this argument, it could be the case that while the financial literature by its very scope considers financial uncertainty to be important, firm employee reading about financial uncertainty is perhaps only a matter of interest unrelated to firm outcomes.[5] Despite this evidence, we find that firm employee reading predicts firm outcomes and financial risk. This can be viewed as a validation of the literature about attention in economics and finance. This literature emphasizes the importance of attention in decision-making, as attention shapes the information that individuals and firms use to make choices (see, e.g., Gabaix, 2014; Peng and Xiong, 2006). Coibion, Gorodnichenko, and Ropele (2020), for example, find that inflation news only affects firm decisions when the firm is primed with information about inflation for a subset of Italian firms.

The remainder of the paper is organized as follows. In section 1, we describe the data. In section 2, we motivate our measure and validate it in section 3. In section 4, we present results on real firm outcomes. Section 5 explores the asset-pricing implications of our measure and section 6 concludes.

[5]While Hassan et al. (2019) shows how discussion of political risk in conference calls between investment analysts and firm executives, it is much less obvious how rank-and-file firm employee reading about financial uncertainty relates to firm financial risk. It is also important to note that while our measure captures executive reading, it is likely that it is only a small component of total reading for public firms.

1 Data

Our proprietary data comes from a company (hereafter "the Consortium") that analyzes content in internet articles published across thousands of media sites (members). Although the Consortium's primary business objective is to supply clients with actionable signals of *intent* to purchase specific business-to-business products and services, the scope and variety of topics covered by their text corpus is broad and diverse. This diversity stems from the Consortium's member pool that spans numerous industries and businesses. Members range from generalist publishers such as The Wall Street Journal, Forbes, and Bloomberg, to more specialized and niche content providers such as Hart (energy), StepStone (private equity), and Quin Street (consumer products). Figure (1) presents a small fraction of the approximately 4,000 publishers that supply the Consortium with its data.

In return for providing raw data, the Consortium's members receive data analytic services that describe how users (i.e., readers) are engaging with the members' published content. This exchange enables the Consortium to view a significant amount of online reading activity; on a typical day they see over 2 *billion* user interactions with member content. These user-article interactions offer valuable insights into the daily reading habits of employees across various firms (e.g., both public and private and for- and non-profit). For each interaction, the Consortium receives several data points, including the URL of the specific online content being read, the user's external IP address, and their cookie data. Leveraging the URLs of each online article, the Consortium employs a state-of-the-art Natural Language Processing (NLP) algorithm to distill the key topics of every article (described below). The Consortium then uses the supplied IP addresses and cookie data to associate users with domains and link those domain-interactions with topics. This process provides us with granular data on the degree to which each firm (i.e., domain) is paying attention to a specific topics on a daily basis.

1.1 Topic Decomposition

The Consortium's key intellectual property is its state-of-the-art machine learning algorithms that helps distills individual news articles into their essential topics. These topics are combinations of words and associations that are learned using a comprehensive set of training corpora and validated exhaustively through human intervention (see, e.g., Gentzkow et al. (2019) for an overview of these textual analysis methods). They come in two varieties. "Specific" topics are created with the express purpose of providing insights to client firms that sell particular products or services. A biotechnology firm, for example, could request information on which companies are researching "RNA sequencing" and "cancer genomics", employing this data to orient its sales or fundraising teams towards prospective interested parties. On the other hand, "general" topics are created to enhance the fit of the Consortium's NLP algorithm by subsuming common variation in reading across users and domains (e.g., users read about politics, vacations, and sports). One would assume that interactions with these "general" topics offer limited insights into a firm's underlying business operations.

Measuring interactions with topics is complicated by the fact that each article is potentially a combination of many topics. A piece covering "RNA sequencing" and "cancer genomics" might, for example, may also reference the "U.S. Food and Drug Administration (FDA)" in passing. To address this issue, the Consortium's NLP algorithm decomposes each article into its constituent topics, providing a proportionality score indicating the significance of each topic in an article. Returning to the example above, the Consortium's algorithm may determine that 50% of the article is related to "RNA sequencing," 45% of the article is related to "cancer genomics," but only 5% of the article is related to the FDA.

After applying its NLP algorithm to the hyper-dimensional dataset of over 2 billion daily *user-article* interactions, the Consortium constructs a lower dimensional dataset of daily *domain-topic* interactions by summing the proportionality scores across all *user-article* interactions at a given domain on each day. On a typical day, this *domain-topic* dataset features roughly 2 million domains and about 6,000 topics.[6] Training the Consortium's NLP algorithm on the set of daily *user-article* interactions is both financial expensive and time consuming. We, therefore, take the resulting *domain-topic* dataset as a fixed part of our analysis, precluding us from augmenting the set of topics in the Consortium's taxonomy to include additional topics and themes.

Figure 2 provides an example of the process. It shows a fictional domain (xyz.com) with three users on 11/17/2018. Each user reads the same Wall Street Journal (WSJ) article, one user also reads an article on microchip.com, and another user also reads an article by Bloomberg. Each

[6] Before processing the *user-article* interaction data, the Consortium applies a number of filters (e.g., bot-detection algorithms) to streamline the data and focus on only meaningful interactions with published content.

publisher feeds this user-article interaction data to the Consortium, who apply their NLP algorithm to each article and determine that the microchip.com article is entirely about CPUs, the WSJ article is a 30%/70% split between inflation and the Federal Open Market Committee (FOMC), and the Bloomberg article is a 20%/80% split between inflation and politics. The Consortium then aggregates these interactions and scores across users and topics to produce the domain-topic interaction data for the day. The scores show that xyz.com primary focus was FOMC news (three users interacted with an article that was 70% about the FOMC, resulting in a score of 210) and the least attention to political news (one user interacted with an article that was 80% about politics, resulting in a score of 80).

This dataset of domain-topic interactions provides us with granular details on the specific set of topics that each firm is paying attention to daily. However, directly using this daily data in our analyses is problematic because reading activity displays a variety of intra-week effects. For example, there are significantly fewer user-article interactions on weekends than on weekdays, and a different composition of topics read on Mondays versus Fridays. Figure OA.2.1 in the Online Appendix highlights this intra-week pattern in reading by showing that on an average Tuesday, 93 unique users per firm (across all roughly 2 million domains) interact with the Consortium's data, while on Fridays, the number falls to 83. We address these intra-week patterns in reading, which are akin to seasonal variation, by aggregating the daily domain-topic interaction data into lower frequencies, such as weekly, monthly, or quarterly. We conduct our analyses using the lower-frequency data.

1.2 Exploring the Data

The Consortium's raw data covers user-level interactions with each of the roughly 6,000 topics in the Consortium's taxonomy. This taxonomy, the members of the Consortium, and thus the number of covered firms, however, has evolved over time. Panel A of Table 1 provides a comprehensive overview of this evolution, presenting annual summary statistics for (a) the number of unique domains covered by the data, (b) the number of topics in the Consortium's taxonomy, and (c) the cross-sectional distribution of employee attention across topics.

There are two main takeaways from Panel A, which covers both public and private firms. First, there is a substantial degree of cross-sectional and time-series heterogeneity in the Consortium's coverage of firms and topics. The number of domains (topics) covered by the Consortium has increased from about 650,000 (2,500) in 2016 to about 2.1 million (7,400) in 2022. This trend reflects the fact that the Consortium has increased its member base over time and the set of topics it covers to satisfy the interests and the demands of new clients. Second, Panel A also shows that the distribution of the number of topics that firms pay attention to in any given week is highly positively skewed. The mean (median) firm pays attention to about 320 (130) topics per week. A skewness of around four suggests that a majority of firms engage with a small subset of the topics.

Panel B repeats the exercise after matching the Consortium's dataset to the CRSP/Compustat universe of firms — the set of domains over which we conduct our empirical analyses in sections 3, 4 and 5.[7] We immediately see that the employees of public firms pay attention to a broader array of topics. The mean (median) number of topics a public firm pays attention to is around 2,800 (2,900). Consequently, the skewness of the firm-topic distribution diminishes to nearly zero among public firms. Comparing the results in Panel A to those in Panel B suggests that the number of topics that a firm pays attention to is inherently correlated with the firm's size. Public firms, which on average are bigger than private firms, unsurprisingly pay attention to more topics at any given point in time.

Table 1 shows that the number of firms that interact with the Consortium's data and the number of topics that the Consortium covers generally increase over time. These changes in the composition of the Consortium's data could potentially distort the time-series dynamics of a firm's attention to any particular topic (e.g., if the Consortium only begins tracking a topic in 2021, then data for that topic will be unavailable for the period ranging from 2016 through 2020). We address this issue by focusing on cross-sectional variation in attention to topics on a given date, ensuring that our results are immune to potential problems associated with (i) the Consortium changing the algorithm used to decompose articles into topics, or (ii) the technological landscape changing (e.g., the possibility that users read more news simply because a faster version of the Wall Street Journal app becomes available on smartphones) over time.

Visualizing the topics. While the Consortium's raw data typically features over 6,000

[7]We only include firms with a CRSP share code of 10, 11, and 12 and a CRSP exchange code of 1, 2, or 3. This confines our analyses to the public equity of firms listed on the NYSE, AMEX, and NASDAQ exchanges. Moreover, we link firms in the CRSP/Compustat universe to the Consortium's dataset via the firm's domain(s). While we remove financial firms and utilities from all empirical tests that involve the CRSP/Compustat universe, we still report attention-related statistics for firms in these industries in Tables 1 and 2 for the purpose of completeness.

individual topics, a careful analysis of the Consortium's taxonomy indicates that some topics are more related to one another than others. To illustrate this point, the Consortium provided us with category labels associated with each topic they follow. For example, the individual topics such as "M&A," "M&A due diligence," and "capital injection" fall under the "corporate finance" category, whereas the "succession planning" and "layoffs" topics fall under the "staff departure" category.

To provide a sense of the types of topics covered by the Consortium's data, we count how many topics are associated with each category on a randomly selected week in the middle of the sample period (11/17/2018). We then present these counts as a word cloud in Figure 3. The category labels in this figure are weighted such that the more prominent categories feature a larger set of related topics. The figure shows that certain categories (e.g., technology and financial services) tend to feature many more topics than other categories (e.g., urban planning). While the distribution of topics across categories is not uniform, it is comforting that it mirrors industries that are a larger proportion of the economy, assuaging concerns of inherent industry biases in the data.

Summary statistics by industry. As the Consortium's coverage of topics is tilted towards the categories of finance, business services, and technology, it is natural to consider both the degree to which we can match firms in the CRSP/Compustat universe to the Consortium's data. Table 2 reports summary statistics along these lines across industry groups. Here, we match each firm to one of 17 2-digit NAICS industry codes. While we report summary statistics for the financial sector and utilities, we follow the convention in the literature and remove these sectors from our empirical tests.

The last row of this table shows that we can successfully match 73% of the 4,394 distinct firms that exist in the CRSP/Compustat universe between 2016 and 2022. The matched firms are typically larger, representing an average of about 81% of the aggregate market capitalization at any given time. Non-price-based measures of firm size, such as the number of employees, show that the matched firms comprise approximately 67% of the total number of employees in the CRSP/Compustat universe. The second to last column also indicates that the employees within an industry interact with about 88% of the topics on average in the Consortium's data.

The preceding rows of this table show the same summary statistics for each of the 17 industry groups. We observe a large degree of heterogeneity: our sample includes only 10 agriculture-related firms but 1,754 manufacturing-related firms. That said, our matched sample still includes the bulk of each industry's market capitalization. For instance, the 10 agricultural firms in our sample represent 99% of the total market capitalization of the agriculture sector. This pattern is generally consistent across industries, with notable exceptions being the real estate and eductation sectors, with only 20% and 41% of their total market capitalization matched, respectively. Lastly, the table reveals varying levels of attention paid to topics by firms in each sector. While firms in education only interact with about 62% of topics in a given week, the employees of the typical retail-based firms interact with over 95% of topics.

Other data-related statistics. For the sake of brevity, Section OA.3 in the Online Appendix provides a battery of additional summary statistics and descriptions of the Consortium's data. We thoroughly examine the distribution of *domain-topic* interactions and show that many topics are only marginally informative about a firm's business line(s). Unsurprisingly, a large proportion of online reading activity of the average firm employee is spent on current events. For example, in the week ending on November 17, 2018, the preponderance of online time was spent on general topics including "South by Southwest," a popular festival and conference related to music and film, and "Call of Duty," a popular online video game which had released a new version shortly before this date. A key insight from this analysis is that rudimentary measures of firm-level attention, such as the total number of interactions scaled by assets or the number of employees, are most likely uninformative about the firm's economic exposures, something we carefully address in the next section.

2 Motivating our Measure of Uncertainty

While economic uncertainty has become a common feature of theoretical frameworks (see, e.g., Bloom, 2009; Colacito, Croce, Liu, and Shaliastovich, 2022, and others) and features prominently in empirical studies (see, e.g., Bali et al., 2017; Ludvigson et al., 2021, and others), finding a suitable proxy for each firms' exposure to uncertainty has proven challenging. For instance, return-based measures, such as the covariance between each firm's excess returns and the VIX index, are noisy estimates of investors', as opposed to firms', perceptions of uncertainty. Likewise, earnings call transcripts only provide occasional insights into the behavior of a subset of key employees in any given firm (see, e.g., Hassan et al., 2019).

Section 2.1 demonstrates using a set of time-series analysis that the average firm's relative attention to uncertainty is highly correlated with well-know proxies for economic uncertainty, including the VIX index, the EPU measure proposed by Baker et al. (2016), and the financial, macroeconomic, and real uncertainty measures of Ludvigson et al. (2021). The central premise of our study, however, is that we can use firm-level attention to uncertainty-related news in near real-time to measure differences in *exposure* to uncertainty. In Section 2.2 we appeal to a well known method in computer science to formulate a data-driven method for differentiating between topics that are informative and uninformative about a firm's economic exposures. We then demonstrate how through this measure we can glean novel economic insights, notwithstanding the average employee's focus on reading about general news and events.

2.1 Attention to Uncertainty-Related Topics

As noted above, firms pay attention to a broad array of topics. To constrain their attention to only uncertainty related topics we measure the extent to which each firm is reading the uncertainty-related topics underlying the EPU index. To do this we first create a sample corpus of uncertainty-related articles using the procedure outlined in Baker et al. (2016). This corpus includes more than 2,500 articles published by the Wall Street Journal (WSJ), the Economist (segmented by six section tabs), Financial Times (FT), Federal Reserve Beige Books (segmented in 12 regions), Federal Reserve Notes, and the Bank of International Settlements. As per the approach of Baker et al. (2016), we only retain articles by these publishers that include the following terms: "*economic*," "*uncertainty*," and either "*Congress*," "*legislation*," "*regulation*," or "*White House*." We then have the Consortium deploy its proprietary NLP algorithm onto this uncertainty-related corpus. The topics that emerge from this exercise are an approximate 600 topic subset of the roughly 6,000 total topics in the Consortium's taxonomy.

Panel A of Figure 4 shows the types of uncertainty-related topics that emerge from this analysis (more prominently read topics are represented as bigger words). Interestingly, the set of popular topics that emerge from the articles underlying the EPU corpus are related to multiple broad categories. In the language of Ludvigson et al. (2021), several prominent topics are related to the real side of the economy (e.g., "Consumer Spending," "Economic Growth," and "Economic Inequality"), while others are related to financial markets (e.g., "Exchange Rate," "Interest Rate," and "Market Volatility"). This goes to show that the there is not necessarily a single facet of economic uncertainty. Rather, the articles underlying the EPU corpus reflect a wide variety of concerns that market participants face.

To illustrate how average reading of uncertainty changes in the time series, we start with a simple of measure of attention to uncertainty at the firm level. We first designate $Total$, Unc and $Other$ as the sets of total topics, EPU-related uncertainty topics, and the remaining set of other topics, respectively,

$$\begin{aligned} &Unc \cup Other = Total \qquad \text{and} \\ &Unc \cap Other = \varnothing. \end{aligned}$$

For each firm, we stack its reading into vectors associated with Unc and $Total$. Each element is the fraction of the firm's employees reading that topic, where topics not read by a firm are given a value of zero. For reasons that will become obvious in the next section we refer to these vectors as a firm's topic-frequency or $\boldsymbol{tf}$ vector. Our measure of attention to uncertainty for firm i at time t is the dot product of the uncertainty versus total vector of reading,

$$RRA_{i,t} = \cos\left(\boldsymbol{tf}_{i,t}^{Unc}, \boldsymbol{tf}_{i,t}^{Total}\right) = \frac{\boldsymbol{tf}_{i,t}^{Unc} \cdot \boldsymbol{tf}_{i,t}^{Total}}{\|\boldsymbol{tf}_{i,t}^{Unc}\| \times \|\boldsymbol{tf}_{i,t}^{Total}\|}, \tag{1}$$

which we refer to as the firm's raw relative attention (RRA) to uncertainty.[8] We focus on *relative* attention because of the highly non-linear relationship between firm size and employees attention to topics. For instance, a large corporation such as Microsoft, with a workforce of over 150,000 employees, will naturally consume more information about uncertainty than a smaller corporation like Malibu Boats, which has approximately 600 employees. However, a smaller, but still large corporation, such as Red Hat, does not necessarily engage with significantly fewer topics than Microsoft, despite being almost a tenth of its size. We provide a graphical illustration of this non-linear relationship by plotting the average length of $\boldsymbol{tf}_{i,t}^{Unc}$ and $\boldsymbol{tf}_{i,t}^{Total}$ versus log firm size in Figure OA.3.5. Due to this non-linear relationship uncertainty reading cannot be compared across firms

[8]We measure the length of each vector by its ℓ^2 (or Euclidean) norm. Additionally, as the measure is a dot product it is bounded between zero and 1. Finally, the $\boldsymbol{tf}^{Unc}$ vector is a member of $\boldsymbol{tf}^{Total}$; as such we commonly refer to $RRA_{i,t}$ as the proportion of uncertainty versus total reading by firm i at time t.

by simply dividing $\boldsymbol{tf}_{i,t}^{Unc}$ by firm size proxies such as the number of employees, assets, or market capitalization. What is also clear in the figure, however, is that both the length of Unc and $Total$ follow similar patterns across firm size; the dot product of these two vectors, however, does not exhibit as stark of a non-linear relationship with firm size.

We next show that during periods of heightened uncertainty, the employees of the average firm tend to read more uncertainty-related news than other business-related news. Consequently, surges in aggregate employee attention to uncertainty-related topics coincide with notable periods of heightened economic uncertainty in our sample period. We establish these facts by aggregating $RRA_{i,t}$ to uncertainty at time t across all firms indexed by $\mathcal{I}_t$,

$$\widetilde{RRA}_t = \frac{\overline{\boldsymbol{tf}}_{i,t}^{Unc} \cdot \overline{\boldsymbol{tf}}_{i,t}^{Total}}{\|\overline{\boldsymbol{tf}}_{i,t}^{Unc}\| \times \|\overline{\boldsymbol{tf}}_{i,t}^{Total}\|}, \tag{2}$$

where $\overline{\boldsymbol{tf}}_{i,t} = \frac{1}{\mathcal{I}_t}\sum_{i=1}^{\mathcal{I}_t} \boldsymbol{tf}_{i,t}$. Figure 5 plots $\widetilde{RRA}_t$ from equation (2) against four standard proxies for uncertainty: the VIX index, the EPU index of Baker et al. (2016), and the financial and macro uncertainty measures of Ludvigson et al. (2021). The primary observation from this figure is that the time-series dynamics of $\widetilde{RRA}_t$ are positively, but not perfectly, correlated with these common proxies for uncertainty. For instance, $\widetilde{RRA}_t$ spikes in late 2019, prior to the onset of the COVID-19 pandemic in early 2020, and tends to rise, but not as much as the other proxies of uncertainty, at the start of the COVID-19-induced recession of March 2020. There seems to be relatively low correlation between our measure, VIX and EPU, and relatively high correlation between our measure and those of Ludvigson et al. (2021). This is unsurprising as there has been substantial work done on how fluctuations in volatility indices can be driven by both changes in uncertainty and risk aversion (see, e.g. Bekaert, Engstrom, and Xu, 2022). Ludvigson et al. (2021), in particular, distinguishes these two sources of spikes in conditional volatility. To quantify and reiterate this higher correlation, in Table 3 we show a correlation matrix across all measures. Correlation coefficients range from about 42% for the VIX and EPU indexes to about 64% for the uncertainty measures of Ludvigson et al. (2021). This strengthens our confidence that variation in a firm's attention to uncertainty-related topics does indeed convey information about the underlying state of the economy.

Our data, however, comports an additional benefit versus traditional text-based measures of risk and uncertainty in that it allows us to distinguish the reading of firms in the cross section.

A series of questions naturally follow: what topics truly separates the attention of firms? Are there certain topics that are more informative and others that are less about uncertainty and the real economy? And do these topics tell us anything economically about firm exposure to these factors? In the next section, we attempt to exploit the richness of the data described above to generate an economically meaningful measure of attention to uncertainty across firms that we will then rigorously test in our empirical results sections.

2.2 Topic frequency, inverse aggregate frequency

We differentiate between informative uncertainty-related topics and less informative topics by adopting a common approach from the field of document retrieval in computer science. Tasks, such as search operations, depend on specific word sequences within a query to extract related documents from a large corpus. Central to this process is the concept of the *tf-idf* (term frequency-inverse document frequency) score, where each word w within a particular document d in the corpus C is assigned a score defined as:

$$\textit{tf-idf}_{d,w,C} = \underbrace{\frac{\text{\# Word in Document}}{\text{Total Words in Document}}}_{\text{Term frequency}} \times \underbrace{\frac{\text{\# Documents in Corpus}}{\text{\#Documents featuring Word}}}_{\text{Inverse document frequency}}. \qquad (3)$$

The *tf-idf* scoring mechanism emphasizes the importance of unique words while downplaying the importance of common ones. For example, prepositions frequently appear in individual documents (i.e., have high *tf*) but are also used widely across documents (have low *idf*). This widespread usage diminishes the usefulness of prepositions in distinguishing between documents in a large corpus of text. Consequently, prepositions generally have low *tf-idf* scores. In essence, the methodology identifies the set words that most differentiate documents.

We build on this idea and propose a novel measure for differentiating between informative and uninformative topics for a firm. We refer to this measure as the *topic frequency-inverse aggregate frequency*, or *tf-iaf*, score. This scoring strategy downweight common topics, such as those associated with current events, that pervade reading across firms, and upweight topics with high readerships within each firm, enabling us to identify topics that distinguish firm reading effectively. To demonstrate the power of this procedure, we apply the *tf-iaf* procedure to non-EPU related (i.e., *Other*) topics, where the effects are visually most stark. Define this score for topic j and firm i at time t as:

$$\begin{aligned} \textit{tf-iaf}_{i,j,t}^{Other} = & \underbrace{\left(\begin{array}{c}\text{Fraction of Employees at Firm } i \\ \text{Interacting with Non-EPU Topic } j \text{ at time } t\end{array}\right)}_{\text{Topic frequency (non-EPU)}} \times \\ & \underbrace{\left(\begin{array}{c}\text{Average Fraction of Employees Across All} \\ \text{Firms Interacting with Non-EPU Topic } j \text{ at time } t\end{array}\right)^{-1}}_{\text{Inverse aggregate frequency (non-EPU)}} \\ \equiv & \; tf_{i,j,t}^{Other} \times iaf_{j,t}^{Other}. \end{aligned} \tag{4}$$

Here, the *tf* component is the fraction of total users within a firm interacting with a topic on a given day. This component is identical to the inputs in equation (1) and effectively places a higher importance on topics with a larger number of interactions. In contrast, the *iaf* component down weights the topics that receive a large number of interactions across all firms.

To illustrate the effect of these *tf-iaf* weights for the non-EPU topics, consider the topic clouds presented in Figure 6. The figures in the top row display the reading for the chemicals manufacturing industry (NAICS 334), whereas the bottom figures in the bottom row reports the reading for the computer and electronics manufacturing industry (NAICS 325). The word clouds in the first column are constructed using the average tf within each industry across the 2 million firm sample for the week ending November 17, 2018. Interestingly, they show virtually no differences in raw reading between the two sectors. Prominent topics include "Live Streaming," "South by Southwest," "Blu-Ray," "United States Secret Service," and "Call of Duty." In the parlance of the *tf-iaf* scores, these topics, however, will also have low inverse aggregate frequency (iaf) scores because of the significant number of interactions across firms. This is because knowing that a firms' employees are reading a common topic, such as "Blu-ray," does little to inform us about whether these employees work in the chemicals or the computer sector.

The second column of Figure 6 demonstrate how these raw word clouds change when we weight each topic by its *tf-iaf* score from equation (4). While current events dominate the raw word clouds, the weighted topics are specific to each industry. For instance, we see that the "Cancer Genomics," "Drug Discovery," and "Angiogenesis Inhibitors" topics are important for firms in the chemicals sector, which includes pharmaceutical firms, whereas the "Cisco ACI," "Remote Desktop Protocol," and "Software-Defined Wide Area Network" topics are important for firms in the computer industry. The success of this procedure stems from the fact that while topics, such as "Cancer Genomics" ("Cisco ACI"), have a large *tf* score within the chemicals (computing) industry, they also have high *iaf* scores because few firms outside the chemical (computing) industry read these topics.

We can apply the same notion of *tf-iaf* weights to the set of topics drawn from the EPU corpus to differentiate between the more and less important uncertainty-related topics. We define the *tf-iaf* score of EPU topic j for firm i at time t as

$$\begin{aligned} \textit{tf-iaf}^{Unc}_{i,j,t} = & \underbrace{\begin{pmatrix} \text{Fraction of Employees at Firm } i \\ \text{Interacting with EPU Topic } j \text{ at time } t \end{pmatrix}}_{\text{Topic frequency (EPU)}} \times \\ & \underbrace{\begin{pmatrix} \text{Average Fraction of Employees Across All} \\ \text{Firms Interacting with EPU Topic } j \text{ at time } t \end{pmatrix}^{-1}}_{\text{Inverse aggregate frequency (EPU)}} \\ \equiv & \; tf^{Unc}_{i,j,t} \times iaf^{Unc}_{j,t}. \end{aligned} \tag{5}$$

Here, the definition of $\textit{tf-iaf}^{Unc}_{i,j,t}$ follows that of $\textit{tf-iaf}^{Other}_{i,j,t}$ from equation (4). The key difference being that the term frequency component is now defined over the set of EPU-related topics versus defined over the set of non-EPU-related topics.

Panel B of Figure 4 shows a similar weighted word clouds, but using only the subset of EPU-related topics for the week ending November 17, 2018. While the raw word cloud in Panel A of Figure 4 shows that many firms read about "Quantitative Easing," "Consumer Spending," and "Economic Growth," these prominent topics are relatively uninformative in distinguishing reading about uncertainty in the cross section of firms. Rather, what truly seems to distinguish between the reading of firms in the cross section are finance-related terms that come in two broad categories: (1) topics such as "credit risk," "exchange rate," and "duration management" that closely relate to firm-level management of risk exposures, and (2) topics such as "Basel II" and "International Accounting Standards Board" that closely relate to regulatory compliance. The fact that finance-related topics are highly related to uncertainty is not only consistent with Ludvigson et al. (2021), but is also surprising in that financial firms are excluded from our analysis, including in constructing these word clouds.

We now introduce our primary measure of interest. We measure firm i's exposure to uncertainty at time t via the degree to which its employees are paying attention to important uncertainty- versus other business-related topics via

$$ARA_{i,t} = \frac{\boldsymbol{tf\text{-}iaf}_{i,t}^{Unc} \cdot \boldsymbol{tf\text{-}iaf}_{i,t}^{Total}}{\|\boldsymbol{tf\text{-}iaf}_{i,t}^{Unc}\| \times \|\boldsymbol{tf\text{-}iaf}_{i,t}^{Total}\|}. \tag{6}$$

Here, $\boldsymbol{tf\text{-}iaf}_{i,t}^{Unc}$ is the vector of *tf-iaf* weights related to the uncertainty-related topics from equation (5), whereas $\boldsymbol{tf\text{-}iaf}_{i,t}^{Total}$ is a vector of *tf-iaf* weights on total reading. The time index t represents attention to each set of topics over a weekly (or lower) frequency to avoid the intra-week effects in reading activity discussed earlier (recall Figure OA.2.1 in the Online Appendix).

The qualifier "important" used to describe the topic weights in equation (6) is critical in distinguishing this measure from that of equation (1). Reading about current uncertainty-related episodes is likely common across agents in the economy and follows the aggregate supply of that news in the economy. This is the straightforward intuition behind both Baker et al. (2016)'s EPU measure and equation (1). In contrast, agents that are more (less) exposed to uncertainty ex-ante will actively try to more (less) aggressively mitigate those risks. Our basic premise is that firms that pay more attention to important uncertainty-related topics (i.e., those that have high *tf-iaf* scores) versus topics associated with other important business activity are more systematically *exposed* to macroeconomic uncertainty.[9] Our hypothesis is closely related to the literature linking uncertainty to both hedging and risk mitigation activity (see, e.g., Brown, 2001; Campello et al., 2011) as well as to regulatory exposure (see, e.g., Kalmenovitz, 2022). Whether this premise is valid or not is an empirical question that we rigorously test in the following sections.

3 Validating the Relative Attention Measure

This section conducts a comprehensive set of empirical exercises to substantiate the informativeness of our relative attention measure in quantifying firm-level exposures to uncertainty. Section 3.1

[9] Intuition would also suggest that exposure to uncertainty, while likely time-varying, is considerable more persistent than the actual arrival of uncertainty-related events. For example, while there is ample evidence of time varying CAPM β, few would doubt that high CAPM exposed stocks tend to remain high β for extended periods of time.

shows that firms with high *tf-iaf* adjusted relative attention scores (i.e., high financial uncertainty) are more exposed to shocks to proxies for aggregate uncertainty discussed in the previous section; specifically, the VIX volatility, Baker et al. (2016) EPU and Ludvigson et al. (2021) uncertainty indices. Sections 3.2 and 3.3 show that firms with high relative attention scores attempt to mitigate their high exposures to uncertainty. These efforts, however, are by their very nature imperfect, resulting in more uncertain firms also having higher implied costs of capital (see, e.g., Brown and Toft, 2002, for a theoretical rational for imperfect mitigation).[10]

3.1 Uncertainty betas

While the previous section shows that $\widetilde{RRA}_t$ has a positive time-series correlation with several proxies for uncertainty employed by the existing literature, we are primarily interested in understanding cross-sectional differences in exposures to uncertainty. Before moving to our formal empirical tests of this notion, we first provide evidence that our firm-level measure of relative attention (i.e., $ARA_{i,t}$ from equation (6)) is capable of distinguishing between firms with high and low exposures to uncertainty.

We do this by first sorting all firms in the cross-section in each week into either tercile or quintile portfolios based on each firm's value of adjusted relative attention from the previous week (i.e., $ARA_{i,t-1}$). We hypothesize that if a firm is highly (sparingly) exposed to economic uncertainty, then changes in the firm's relative attention to uncertainty should coincide strongly (weakly) with changes in these common proxies for uncertainty. That is, firms with high (low) values of $ARA_{i,t}$ in the recent past are more likely to read relatively more (less) about uncertainty when a proxy such as the VIX or EPU index rises (falls). We test this hypothesis by regressing the relative attention of each uncertainty-sorted portfolio on each of the common uncertainty proxies as follows:

$$\widetilde{RRA}_{p,t} = \delta_p + \beta_p^m U_t^m + \epsilon_{p,t}, \tag{7}$$

where $\widetilde{RRA}_{p,t}$ is the average relative attention of portfolio p to uncertainty at time t. Here, $\widetilde{RRA}_{p,t}$ is defined as per equation (2) with the exception that the summation is taken over the subset of firms that belong to portfolio p. U_t^m denotes the time t realization of uncertainty proxy m. The

[10] In section OA.4 in the Online Appendix we provide a stylized model highlighting the key reasons behind why firms are incapable of perfectly hedging all risks.

key variable of interest is β_p^m. This slope coefficient captures the sensitivity of each portfolio's raw relative attention to uncertainty to changes in the uncertainty proxy of interest. We run regression (7) on each tercile or quintile portfolio as well as on the difference of the extreme portfolios (i.e., either 3 minus 1 or 5 minus 1) and assess the statistical significance of β_p^m using Newey-West adjusted standard errors. It's important to note that while the VIX and EPU are available on a weekly basis—the highest frequency at which our data is available—the Ludvigson et al. (2021) measures are only available monthly.

Table 4 reports the results for tercile-sorted portfolios (Panel A) and quintile-sorted portfolios (Panel B). The main takeaway from this table is that the raw uncertainty reading of firms that have high values of $ARA_{i,t}$ in the past are more sensitive to future fluctuations in uncertainty. In other words, the set of firms that we label as having high exposure to economic uncertainty (due to their high values of $ARA_{i,t}$) also shift their reading towards relatively more uncertainty-related news in the periods when each of these common proxies for uncertainty rises. This positive spread in uncertainty betas between the high and low ARA portfolios is present for both the tercile- and quintile-sorted portfolios. The uncertainty spread is statistically significant at the 5% level for all portfolios in Panels A and B. Of additional note is that for many of the uncertainty proxies, the $ARA_{i,t}$ sorted portfolios are monotonic in exposure as we move from the first to third or fifth portfolios. Taken together, these results show that even though our firm-level measure of adjusted relative attention from Section 2.2 is motivated by cross-sectional differences in reading, firms with high attention to important uncertainty topics tend to have higher βs (i.e., more exposure) to fluctuations in traditional proxies of uncertainty considered by the literature.[11]

3.2 Risk mitigation and compliance

Section 2.2 shows that a consequence of applying the *tf-iaf* weights from equation (5) to cross-sectionally differentiate between topics that are informative and uninformative about uncertainty exposures is that topics related to the mitigation of risk and regulatory compliance tend to carry higher weights. Recalling Figure 4, we saw that the *tf-iaf* procedure shifts uncertainty topic weights from topics such as "Quantitative Easing" (QE) and "Consumer Spending" to topics such as "Dura-

[11] A natural question is whether sorting on $RRA_{i,t}$ itself generates a similarly stark difference in portfolio exposures to uncertainty. Table OA.2.2 in the Online Appendix show that results are weaker, not statistically significant, and highly non-monotonic.

tion Management," "Cost of Capital," "Basel II" and "International Accounting Standards Board." This suggests that while we do not learn much about a firm's exposure to uncertainty if its employees are reading about QE and consumer spending, we may be able to infer its exposure if they are reading relatively more than other firms about either methods to mitigate risk or approaches to comply with regulatory authorities.

Hedging. To show this fact more rigorously, we appeal to the extensive literature that examines firm hedging incentives. We follow the methodology developed in Campello et al. (2011) to see how a firm's hedging activity correlates with $ARA_{i,t}$ over the past year. Specifically, we count the number of times each of the following uncertainty-related keywords are mentioned in a given firm's 10-K each year: "derivatives," "hedge," "financial instrument," "swap," "market risk," "expos," "futures," "forward contract," "forward exchange," "option contract," "risk management," and "notional." The total occurrence of these hedging-related words is then divided by the total word count of each 10-K to derive a firm-year measure of hedging intensity.

As Campello et al. (2011) note, hedging activity plays little to no role for a large fraction of firms. Consistent with this finding, we note that around 25% of firms mentioned four or fewer hedging-related words in their 10-Ks. A small number of firms, however, mention many of these words in their annual reports. To minimize the effects of outliers, we define our firm-level measure of hedging using an indicator variable that is denoted by $\mathbb{I}^{\text{Hedge}}_{i,t}$. This variable takes on a value of one for firm-year observations that have a heading intensity measure that is above the cross-sectional median measure, and zero otherwise. We use this indicator variable as our dependent variable of interest when examining the relation between hedging intensity and a firm's *tf-iaf* adjusted relative attention to uncertainty.

We formally examine this relation by estimating the following regression:

$$\mathbb{I}^{\text{hedge}}_{i,k,t} = \psi_{k,t} + \beta ARA_{i,k,t-1} + \boldsymbol{Z'}_{i,k,t-1}\boldsymbol{\gamma} + \boldsymbol{U'}_{i,k,t-1}\boldsymbol{\nu} + \epsilon_{i,k,t}, \tag{8}$$

where $ARA_{i,k,t-1}$ is the average value of the relative attention measure of firm i in industry k over the year prior to the firm's fiscal year-end. We average over this time period as 10-Ks reflect firm activity over the entire firm year. Additionally, as discussed in greater detail below, we map ARA to be between 0 and 1 to both remove the effect of outliers and for ease of explaining our measure's economic content. The vector $\boldsymbol{Z'}$ contains a host of firm-level controls, including those from Leary and Roberts (2014) and the measure of financial constraints from Whited and Wu (2006), while the vector $\boldsymbol{U'}$ contains three uncertainty-related controls: the exposure of each firm to fluctuations in the VIX index, and the political and non-political risk measures from Hassan et al. (2019). The inclusion of $\boldsymbol{U'}$ helps us to ascertain whether our relative attention measure is conveying any information about a firm's exposure to uncertainty that is not already captured by common proxies in the literature. Finally, $\psi_{k,t}$ reflect industry-by-time fixed effects that subsume unobserved heterogeneity in hedging activity that could differ between industries and over time. We assess the statistical significance of the key variable of interest, β, by clustering standard errors by both firm and year.

Panel A of Table 5 presents the results of estimating equation (8). The first column displays the regression without controls and fixed effects. The coefficient on $ARA_{i,t-1}$ indicates that as a firm moves from the 0^{th} to 100^{th} percentile of the ARA measure, the probability of having an above median mention of hedging-related words increases by approximately 30% (t-statistic of 4). In the second and third columns, we add date and industry fixed effects, respectively, and show that the magnitude and statistical significance of the aforementioned effect remain essentially unchanged. This shows that firms that are paying relatively more attention to uncertainty are indeed more likely to mitigate the risks they face through their hedging activities.

Column (4) of Table 5 augments Column (3) by including an industry-by-time fixed effect in place of industry and time fixed effects and our battery of controls. While the new fixed effects specification subsumes unobserved heterogeneity in hedging activity between industries that can change over time, the slope coefficient on $ARA_{i,t-1}$ remains positive and significant at the 1% level (t-statistic of 3.6). Likewise, the slope coefficient on ARA in Column (5) also remains positive and significant when we further augment Column (4) by including a vector of uncertainty-related controls. The fact that we continue to find a positive relationship between our measure of a firm's exposure to uncertainty and hedging activity, even after we control for a variety of common proxies for firm-level uncertainty, indicates that $ARA_{i,t}$ captures elements of firms' exposure to economic uncertainty that are not simply related to industry affiliation, time effects, how a firm's returns are exposed to fluctuations in the VIX index, or what a firm mentions during its earnings calls (i.e., the information set used in Hassan et al., 2019).

Compliance. While the preponderance of important topics in Panel B of Figure 4 are related to risk mitigation, a number of topics are also related to compliance. This implies that some of the most important topics for differentiating between firms with high and low exposures to uncertainty depend on the extent to which these firms must comply with various laws and regulations. We formalize this connection by examining the relationship between our measure of relative attention to uncertainty and the firm-level regulatory intensity measured in firm-level estimates of dollar expended to satisfy their regulatory burden (see Kalmenovitz, 2022, for procedure to extract estimates from regulatory filings). Specifically, we repeat regression (8) after replacing the left-hand side variable with each firm's regulatory intensity and report the results of these regressions in Panel B of Table 5.

As each column in Panel B shows, the relation between a firm's relative attention to uncertainty and its degree of regulatory intensity is positive and mostly statistically significant. The exception is specification (4); the addition of the Whited and Wu (2006) financial constraint control in particular reduces the statistical significance and magnitude on our coefficient of interest. In Column (5), we present our most comprehensive specification that includes industry-by-time fixed effects, the firm-level controls, and the vector of uncertainty-related controls. We find that firms that have the highest values of $ARA_{i,t-1}$ are required to comply with costlier regulation than firms with the lowest values of $ARA_{i,t-1}$. This difference in regulatory intensity is significant at the 10% level (t-statistic of 1.84). The preceding columns show similar if not stronger results in specifications that include other combinations of fixed effects and controls. These regressions were done at a quarterly frequency as our controls are available at this frequency. Overall, this confirms that firms that spend more time reading about uncertainty-related news tend to be subject to more regulatory oversight.

3.3 Cost of Capital Regressions

Finally, we examine how our measure of firm-level uncertainty covaries with firm-level discount rates. The basic premise of this test is that if economic uncertainty reflects a risk that firms are deferentially exposed to, then firms with higher exposures should not only spend relatively more time reading about uncertainty (i.e., have high values of $ARA_{i,t}$) but should also have higher costs of capital to reflect this increased risk. To shed light on this, we follow the standard approach of

Gebhardt, Lee, and Swaminathan (2001) to infer the cost of equity capital (henceforth, ICC) from equity analysts' earnings forecasts. Section OA.1 of the Online Appendix provides specific details on how we construct this measure. Given our data's short time period, we use implied costs of capital to minimize the effects of noise in realized returns. Nonetheless, table OA.4.4 in the Online Appendix shows that we obtain similar takeaways when we use realized returns instead.

To answer the question of how a firm's cost of capital is related to firm uncertainty, we regress the firm's quarterly cost of capital, $ICC_{i,t}$, on each firm's lagged measure of *tf-iaf* adjusted relative attention using the same regression specification as equation (8), but replacing the left-hand side variable with $ICC_{i,t}$. Here, we estimate the regression using data from the end of each earnings announcement month because it is around this point in time that analysts make the largest adjustments to each firm's earnings estimate for the coming quarters and year. Moreover, much of the fundamental accounting data required to estimate a firm's ICC is only released quarterly.

We present the results in Table 6. The first column indicates a strong positive relationship between a firm's relative attention to uncertainty and ICC. Columns (2) and (3) add combinations of date and industry fixed effects to the regressions and show that the results remain economically and statistically significant at the 1% level even after accounting for unconditional differences in ICC across dates and industries, respectively. The magnitude of the point estimates in Columns (1) to (3) suggests that, on average, a firm moving from the lowest value of ARA to the highest value of ARA sees its cost of capital increase by 150 basis points. While the magnitude of this effect falls when we consider industry-by-time fixed effects and additional controls in Columns (4) and (5), the basic fact remains the same: firms that pay more attention towards uncertainty-related topics have higher costs of capital than those that pay more attention to other business-related news.

4 The Real Effects of Attention to Uncertainty

The previous section shows that firms that pay relative more attention to uncertainty-related topics have higher costs of capital. In this section, we ask if variation in firm attention to uncertainty-related topics is similarly related to real firm outcomes, such as investments, sales, and employment growth rates. We find that firms that pay *relatively* more attention to important uncertainty-related topics invest less in capital, have lower sales growth rates, and have lower employment growth rates. That is, firms that allocate a relatively high proportion of their attention to uncertainty-related topics not only have higher costs of capital but implement contractionary corporate policies.

4.1 Firm-level outcomes and attention to uncertainty

We establish this novel fact via panel regression of the form

$$y_{i,k,t} = \psi_{k,t} + \beta ARA_{i,t-1} + \boldsymbol{Z}'_{i,t-1}\boldsymbol{\gamma} + \boldsymbol{U}'_{i,t-1}\boldsymbol{\nu} + \varepsilon_{i,t}, \tag{9}$$

where $y_{i,k,t}$ denotes a real outcome for firm i in industry k at time t (e.g., the firm's investment or hiring rate). $\psi_{k,t}$ denotes an industry-by-time fixed effect that subsumes common shocks each industry may face at a given point in time (e.g., the possibility that investment opportunities differ between technology and manufacturing firms at the onset of the COVID-19 pandemic). However, certain specifications also include time and industry fixed effects separately. $\boldsymbol{Z}_{i,t}$ is a vector of time-varying and firm-specific controls that capture aspects of firm i's economic environment that are known determinants of its decision to invest. Following Leary and Roberts (2014) this vector includes firm size, Tobin's q, profitability, leverage, and asset tangibility. We also add the firm-level measure of Whited and Wu (2006) to this set. Finally, $\boldsymbol{U}'_{i,t-1}$ is a vector of firm-specific controls that contains various proxies for the firm's exposure to uncertainty. This vector includes the exposure of each firm's stock returns to fluctuations in the VIX index and both the political and non-political risk measures from Hassan et al. (2019). Section OA.1 in the Online Appendix provides details on the construction of these control variables, Table OA.2.1 in the Online Appendix reports summary statistics. We double cluster standard errors by firm and time.

The key variable of interest in equation (9) is $ARA_{i,t-1}$ from equation (6). The point estimate β tells us how a firm's relative attention is related to its real outcomes, conditional upon the comprehensive set of fixed effects and control variables included in equation (9). For the purpose of estimating this equation, we again apply a rank transformation to $ARA_{i,t-1}$ so that this measure of relative attention lies within the unit interval.[12] The benefit of applying this transformation is

[12]Specifically, if $ARA_{i,t-1}$ denotes the raw relative attention for firm i between $t-1$ and t, then its rank transformation is $F(ARA_{i,t-1}) = \text{Rank}\left(C_{i,t-1}\right)/(N_t+1)$, where N_t is the number of firms at time t, $\text{Rank}\left(\min_{i=1,\dots,N_t} C_{i,t-1}\right) = 1$ and $\text{Rank}\left(\max_{i=1,\dots,N_t} C_{i,t-1}\right) = N_t$. This transformation implies that the α-quantile of $F(ARA_{i,t-1})$ is α. For

twofold. First, the transformed measure of relative attention is less sensitive to the presence of outliers. Second, this transformation allows us to interpret the estimated value of β in equation (9) as the difference in firm-level outcomes between a firm paying the least versus the most attention to *tf-iaf* adjusted uncertainty-related topics.

Table 7 reports the results of estimating equation (9) and establishes our novel fact that a higher degree of attention to uncertainty-related topics is associated with a *contraction* in the average firm's future investment and sales. Importantly, this negative association between a firm's attention and real outcomes holds regardless of whether we include industry-by-time fixed effects, thereby capturing differences in relative attention within each industry, the comprehensive set of firm-level controls proposed by Leary and Roberts (2014), the financial constraint measure of Whited and Wu (2006) or a wide variety of uncertainty-related controls that try to capture the firm's exposure to uncertainty. All in all, a firm's relative attention to uncertainty remains highly informative about its real outcomes even after we control for these potentially confounding effects and proxies.

Specifically, Column (1) of the table, which features no fixed effects or control variables, shows that increases in relative attention to uncertainty are associated with a 3% decline in one-quarter ahead investment (t-statistic of -3), a 20% decline in one-quarter ahead sales (t-statistic of -7), and a 20% decline in one-year ahead employment growth (t-statistic of -4).[13] Adding time-fixed effects to the baseline specification in Column (2) leaves these results unchanged and eliminates the concern that the results are driven by a limited number of periods in which all firms cut investment simultaneously (e.g., in the onset of the COVID-19 pandemic in March 2020). Similarly, the regression results in Column (3) shows that even by augmenting Column (2) with industry fixed effects and thereby accounting for fixed differences in investment opportunities across industries, the strong negative relation between uncertainty-related news and investment rates persists.

Columns (4) and (5) present the results of our most comprehensive empirical specifications that include industry-by-time fixed effects, thereby controlling for time-varying differences in investment opportunities across industries (e.g., the possibility that the average technology firm benefited from

notational simplicity, we refer to the rank transformed value of the relative attention measure as $ARA_{i,t-1}$.

[13] Table OA.2.3 in the Online Appendix considers which components of total assets firms may adjust in response to uncertainty and shows that both short-term (i.e., inventory) and long-term (i.e., property, plant, and equipment) assets fall as the relative amount of attention to uncertainty rises.

the onset of the COVID-19 pandemic while the average durable goods retailer suffered). Both columns also feature the large set of firm-level control variables discussed earlier. Column (5) also includes the vector of uncertainty-related controls (i.e., $\boldsymbol{U}'_{i,t-1}$ from equation (9)). Controlling for this vector of alternative uncertainty-related proxies allows us to gauge whether our relative attention measure is providing us with novel information about each firms' exposure to fluctuations in uncertainty or simply providing an alternative means of capturing what is already reflected by a firm's exposure to fluctuations in uncertainty, as measured by the VIX index, and the proxies for political and non-political risk elicited from earnings call transcripts by Hassan et al. (2019).

The results in Column (4) indicate that the robust and negative association between attention to uncertainty-related topics and each of investment, sales growth, and employment growth continues to persist even after we account for both (i) the unobservable heterogeneity in these real effects across industries in time and (ii) the battery of firm-specific controls. Notably, at any given point in time, the average firm in each industry that is allocating most of its relative attention to uncertainty has an investment rate that is 1.32% per quarter less the investment rate of the firm that is paying the least amount of attention to uncertainty. This effect is statistically significant at the 5% level. Likewise, firms paying the most attention to uncertainty have sales growth rates (employment growth rates) that are 5.30% (5.84%) lower than their peers. Most importantly, the results in Column (5) show that the takeaways from Column (4) are not subsumed when we also control for a variety of alternative uncertainty-related proxies, including the political and non-political risk measures proposed by Hassan et al. (2019). This goes to show that understanding a firm's allocation of attention towards uncertainty-related news is highly informative about the real actions the firm will implement in the future.

5 The Risk-return Tradeoff

This section explores the asset-pricing implications of firms' attention to economic uncertainty. Given our very short sample period (i.e., 2016-2022), we focus our analysis on the relationship between ARA and several firm characteristics that the prior literature has related to firm-level risks: market beta, market capitalization, book-to-market, gross profit, and asset growth. We focus on these characteristics as they form the basis of the factors underlying the Fama and French

(2015) five-factor model. We find that while firms that pay relatively more attention to financial uncertainty-related topics are economically and statistically related to these characteristics, there is still substantial unexplained variation in our measure, suggesting that it brings new information to standard empirical asset pricing models.

5.1 Relation to standard characteristics

To show the relation between *ARA* and these firm-level characteristics, we first sort stocks in either three or five *ARA*-ranked portfolios on portfolio formation dates (i.e., calendar quarter-end t) and then calculate the value-weighted average characteristic of each portfolio. Consistent with information revelation on accounting release dates, we update a firm's *ARA* at the month-end of its respective announcement date. Table 8 reports the results of this analysis. Across both the tercile and quintile sorts in Panels A and B, we find that high *ARA* stocks generally load correctly on the standard characteristics. For instance, the high *ARA* portfolio in Panel A has an average gross profitability estimate of 0.26, which is higher than that of the low *ARA* portfolio, which is 0.17. The 0.09 difference in these characteristics is significant at the 1% level (t-statistic of 3.49). The difference between the high and low *ARA* portfolio is positive for both market beta and book-to-market but negative for asset growth, as expected. The differences in asset growth are consistent with the panel regressions from Section 4. The outlying characteristic is the difference in market capitalization, which is positive even though the size anomaly predicts a negative relation with returns.[14] The takeaways remain unchanged in Panel B, which reports results for quintile-sorted portfolios. The strong relationship between *ARA* and these characteristics bolsters our confidence that our sorting variable is strongly related to risk in general.[15]

5.2 Variance Decomposition

One concern given the discussion in the previous section, is that our measure is a reflection of exposure to risk, but *not*, specifically, to uncertainty (see, e.g., Dew-Becker, Giglio, Le, and Rodriguez, 2017, for discussion). Despite the short time series of our dataset, we attempt to address

[14] Although one may be worried that our *ARA* measure does not line up correctly with the size characteristic, it is worth pointing out that the size factor has had a close to zero return over the past two decades.

[15] The strong relation, e.g., may allow us to extrapolate the predictive properties of *ARA* via its relation to the other characteristics into the past.

this concern in two ways. First, in Table 9, we perform a simple variance decomposition of ARA; specifically, the table presents the marginal R^2 when adding various fixed effects or contemporaneous variables to a regression with ARA as the dependent variable. Mimicking our empirical specification in Sections 3 and 4, we first see how a set of fixed effects relate to ARA. As noted earlier, to minimize effects from outliers, we map ARA to be between 0 and 1 on a weekly basis; as such, we ignore the date fixed effects specification used in our earlier regressions. Depending on the granularity of NAICS industry classification, Table 9 shows that sector fixed effects absorb at most 6.5% of ARA's weekly variation. This increases by approximately 3.0% when enhancing the fixed effects specification to include sector-by-date effects. That leaves more than 90% of ARA's variation to be firm-specific. We then add firm fixed effects to the regression. While anywhere between 30-35% of ARA's variation is permanent across firms from 2016-2022, this still leaves almost 60% of ARA's variation to be firm-date specific.

Returning to the characteristic sorts above, we add the same five characteristics to the regression.[16] As time-varying measures, it is possible that information in these common characteristics tells us significantly more about ARA; as Columns (1) and (2) indicate, however, these characteristics capture very little additional variation. In Column (3), we show the same regression without any fixed effects. At most, the firm-level characteristics capture around 12.0% of the variation in ARA. This high level of firm-by-time specific variation strongly suggests that our measure has economic content beyond traditional risk factors considered in the literature.

Second, in Table OA.4.4, we provide additional support using equity market returns. Specifically, we find that firms in the high quintile ARA portfolio tend to have higher average returns than low quintile ARA firms over our sample period. In keeping with the above approach, we update the relative attention sorting variable at a quarterly frequency but allow returns to vary weekly. In Section 3.3, we statistically showed that high ARA firms have higher analyst implied costs of equity capital; as we move from the lowest to highest ARA firm, the ICC increases from anywhere from 1.5-2.0%. In Column (1) of Table OA.4.4, we see that the high-minus-low (HML) ARA portfolio generates a return of about 14 bps per week, which is only statistically significant at the 20% level. Columns (2) and (3) then show that this return is not explained by either the CAPM

[16]Before adding these variables to the regression, and to be consistent with the dependent variable, we map each characteristic to be between 0 and 1 on a weekly basis.

or the Fama and French (2018) three-factor model. The alpha with respect to each of these models is about 16-18 basis points per week and significant at the 10% level and 5% level, respectively. We note, however, that the alpha drops to 6 bps per week in Columns (4) and (5) when we consider the Fama and French (2015) five-factor plus momentum model.

While these latter results are particularly statistically weak, as one would expect within only six years of data, the annualized returns of more than 3.0% overlap closely with the predictive relationship between ARA and ICC described above. Taken together, we conclude that high ARA firms, which we have shown are more exposed to economic uncertainty, do indeed generate higher average returns than low ARA firms. It is important to note, however, that we cannot rule out the alternative hypothesis that high ARA firms are simply mispriced.

6 Conclusion

This paper uses a novel and ultra-high dimensional dataset that captures the daily internet news reading of the employees of public firms to ask three questions: (1) Do individuals or organizations actually pay any attention to uncertainty-related news?; (2) If so, does the type of news they read inform us about their exposure to uncertainty?; and (3) What plausible actions do these agents take to mitigate their risk?

Our findings suggest an affirmative answer to the first two questions. First, aggregate firm attention to uncertainty versus other business-related news have a high correlation with standard measures of economic uncertainty from the literature such as the VIX volatility, Baker et al. (2016) EPU, and Ludvigson et al. (2021) indices. It indeed seems true that firm employees do pay attention to uncertainty. Second, using a procedure similar to that of *tf-idf* scores used in information retrieval, we highlight the types of topics that most distinguish reading between firms in the cross section. We find that these topics tend to be closely related to financial concerns such as hedging activity and regulatory compliance. We then empirically verify that it is higher attention to these subsets of topics that separate more from less uncertain firms. Specifically, firms that read more about these topics have significantly higher uncertainty-reading βs to the same set of uncertainty proxies listed above. Our measure of exposure to uncertainty is thus the relative degree to which firm-employees are paying attention to this subset of topics.

This finding linking the consumption of certain types of news to a firm's exposure to uncertainty dovetails into the third question. First, reading about hedging and compliance pre-supposes that uncertain firms are more involved in these activities. Using a proxy of hedging intensity using information from a firm's own 10-K, we link our measure directly to future risk-management activity (see Campello et al., 2011, for further descriptions of the measure). Second, we find that firms that read the most about uncertainty have between a 2-10% higher regulatory burden than those that read the least using the measures of Kalmenovitz (2022). Third, we find that firms that read the most about uncertainty in the cross-section have implied cost of capital that are roughly 2% higher than firms that read the least. These findings are robust to including firm-by-date fixed effects, a battery of controls from Leary and Roberts (2014), and the firm-level uncertainty proxies from Hassan et al. (2019).

Finally, we link these higher cost of capital to lower investment in capital and hiring. Specifically, the firms with the highest relative attention measures invest about 4.50% (2.0%) per quarter less in physical capital (inventories) and have hiring growth rates that are 5% per annum lower than firms with the lowest attention to uncertainty. We establish these results by estimating firm-level panel regressions that control for a comprehensive set of observable firm characteristics and a variety of fixed effects that account for unobservable differences between industries and across time.

Our results are important because we directly link business-relevant attention to firm risk for a vast cross-section of firms, including almost all public firms. This stands in contrast to many papers in this area that rely on surveys and potentially a non-representative sample of firms (e.g., Coibion et al. (2020)) to explore these questions. Our paper, directly connecting firm attention to firm expectations, also complements a recent literature that uses economic models with learning to generate inferred measures of firm expectations and suggestive relationships between firm observable outcomes (see, e.g., Kim, Kuehn, and Li, 2021). While we focus on a firm's attention to uncertainty in this paper, asking whether firms pay attention to other sources of risk, such as climate risk and regulatory risk, provides interesting avenues for future research.

References

Alfaro, Ivan, Nicholas Bloom, and Xiaoji Lin, 2018, The finance uncertainty multiplier, Technical report, National Bureau of Economic Research.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Brent Meyer, and Nicholas Parker, 2022, Surveying business uncertainty, *Journal of Econometrics* 231, 282–303, Annals Issue: Subjective Expectations & Probabilities in Economics.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, 2016, Measuring economic policy uncertainty, *The quarterly journal of economics* 131, 1593–1636.

Bali, Turan G, Stephen J Brown, and Yi Tang, 2017, Is economic uncertainty priced in the cross-section of stock returns?, *Journal of Financial Economics* 126, 471–489.

Bartram, Söhnke M, Gregory W Brown, and Jennifer Conrad, 2011, The effects of derivatives on firm risk and value, *Journal of Financial and Quantitative Analysis* 46, 967–999.

Bekaert, Geert, Eric C Engstrom, and Nancy R Xu, 2022, The time variation in risk appetite and uncertainty, *Management Science* 68, 3975–4004.

Bernanke, Ben S, 1983, Irreversibility, uncertainty, and cyclical investment, *The quarterly journal of economics* 98, 85–106.

Bloom, Nicholas, 2009, The impact of uncertainty shocks, *econometrica* 77, 623–685.

Bloom, Nicholas, 2014, Fluctuations in uncertainty, *Journal of Economic Perspectives* 28, 153–76.

Brown, Gregory W, 2001, Managing foreign exchange risk with derivatives, *Journal of Financial Economics* 60, 401–448.

Brown, Gregory W, and Klaus Bjerre Toft, 2002, How firms should hedge, *The review of financial studies* 15, 1283–1324.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* 1, 195–228.

Campello, Murillo, Chen Lin, Yue Ma, and Hong Zou, 2011, The real and financial implications of corporate hedging, *The journal of finance* 66, 1615–1647.

Chinco, Alex, Samuel M Hartzmark, and Abigail B Sussman, 2022, A new test of risk factor relevance, *The Journal of Finance* 77, 2183–2238.

Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele, 2020, Inflation expectations and firm decisions: New causal evidence, *The Quarterly Journal of Economics* 135, 165–219.

Colacito, Riccardo, Mariano M Croce, Yang Liu, and Ivan Shaliastovich, 2022, Volatility risk pass-through, *The Review of Financial Studies* 35, 2345–2385.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2017, The price of variance risk, *Journal of Financial Economics* 123, 225–250.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1 – 22.

Fama, Eugene F, and Kenneth R French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Keith Kuester, and Juan Rubio-Ramírez, 2015, Fiscal volatility shocks and economic activity, *American Economic Review* 105, 3352–3384.

Gabaix, Xavier, 2014, A sparsity-based model of bounded rationality, *The Quarterly Journal of Economics* 129, 1661–1710.

Gebhardt, William R, Charles MC Lee, and Bhaskaran Swaminathan, 2001, Toward an implied cost of capital, *Journal of accounting research* 39, 135–176.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–74.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun, 2019, Firm-level political risk: Measurement and effects, *The Quarterly Journal of Economics* 134, 2135–2202.

Hou, Kewei, Mathijs A Van Dijk, and Yinglei Zhang, 2012, The implied cost of capital: A new approach, *Journal of Accounting and Economics* 53, 504–526.

Jurado, Kyle, Sydney C Ludvigson, and Serena Ng, 2015, Measuring uncertainty, *American Economic Review* 105, 1177–1216.

Kalmenovitz, Joseph, 2022, Regulatory intensity and firm-specific exposure, *Forthcoming,The Review of Financial Studies* .

Ke, Zheng Tracy, Bryan T Kelly, and Dacheng Xiu, 2019, Predicting returns with text data, Technical report, National Bureau of Economic Research.

Kim, Yongjin, Lars-Alexander Kuehn, and Kai Li, 2021, Learning about the consumption risk exposure of firms, *Available at SSRN 3897353* .

Leary, Mark T, and Michael R Roberts, 2014, Do peer firms affect corporate financial policy?, *The Journal of Finance* 69, 139–178.

Ludvigson, Sydney C, Sai Ma, and Serena Ng, 2021, Uncertainty and business cycles: exogenous impulse or endogenous response?, *American Economic Journal: Macroeconomics* 13, 369–410.

Mullainathan, Sendhil, and Andrei Shleifer, 2005, The market for news, *American economic review* 95, 1031–1053.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation, *Econometrica* 55, 703–708.

Pástor, L'uboš, Meenakshi Sinha, and Bhaskaran Swaminathan, 2008, Estimating the intertemporal risk–return tradeoff using the implied cost of capital, *The Journal of Finance* 63, 2859–2897.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Stewart, Craig A, Von Welch, Beth Plale, Geoffrey Fox, Marlon Pierce, and Thomas Sterling, 2017, Indiana university pervasive technology institute .

Stulz, René M, 1984, Optimal hedging policies, *Journal of Financial and Quantitative analysis* 19, 127–140.

Whited, Toni M, and Guojun Wu, 2006, Financial constraints risk, *The review of financial studies* 19, 531–559.

Zingales, Luigi, 2015, Presidential address: Does finance benefit society?, *The Journal of Finance* 70, 1327–1363.

Tables and Figures

Table 1: Summary Statistics
This table presents summary statistics for the number of firms (*domains*) covered by the data, the number of topics in the Consortium's taxonomy, and the cross-sectional distribution of firm-topic interactions for each year in our sample. Panel A reports results for all firms, while Panel B focuses on the CRSP-COMPUSTAT universe of firms. The cross-sectional distribution of firm-topic interactions is characterized by the mean, median, standard deviation (*SD*), and skewness, estimated using data from the last week of each year. The sample period is from June 2016 to July 2022.

(a) All firms

			Cross-sectional Firm-Topic			
Year	Firms	Topics	Mean	Median	SD	Skew
2016	651599	2407	204.27	111.00	274.88	3.02
2017	1935696	3227	477.77	256.00	582.79	2.02
2018	2051067	4804	444.62	176.00	687.69	2.95
2019	1713267	5500	262.84	76.00	567.97	4.83
2020	1960832	5869	306.06	105.00	587.48	4.49
2021	2053800	7392	301.77	86.00	642.63	5.01
2022	2124557	7395	295.76	83.00	630.07	4.98

(b) Public firms

			Cross-sectional Firm-Topic			
Year	Firms	Topics	Mean	Median	SD	Skew
2016	2609	2407	992.04	957.00	653.37	0.17
2017	2749	3227	2406.45	2864.00	964.07	-1.13
2018	2755	4804	3229.05	3820.00	1578.06	-0.71
2019	2786	5500	2933.25	3055.50	1962.48	-0.12
2020	2877	5869	3110.18	3185.00	2026.29	-0.10
2021	3181	7392	3426.36	3212.00	2505.42	0.13
2022	3267	7395	3410.00	3204.00	2465.62	0.15

Table 2: Summary Statistics and Characteristics by Industry

This table reports average industry characteristics across 17 North American Industry Classification (NAICS) 2-digit industries, providing insights into the relationship between industry characteristics and the Consortium's data. For each industry, we report summary statistics for all public firms (*Public*) and the fraction of public firms that we match to the Consortium's data (*Match (%)*). The statistics include the number of firms in each industry (*Firms*), the market capitalization of each industry (*Market Cap*), and the number of employees in each industry (*Employees*). Additionally, we report the fraction of topics in the Consortium's data that firms in each industry interact with (*Topics (%)*) and the average value of our relative attention to uncertainty measure (*Attention (%)*), as defined in Section 1. The topic- and attention-related statistics are first calculated at the firm level across all years and then averaged across all firms in that industry. The definition of each NAICS is available at the US Census website, `https://www.census.gov/naics/`.

	Firms		Market Cap		Employees		Topics	Attention
Industry	Public	Match(%)	Public	Match(%)	Public	Match(%)	(%)	(%)
Education	29	48.00	1364572	41.26	8147	40.18	62.18	53.66
Information	456	69.46	14932460	67.21	13400	43.46	90.02	54.78
Technical Services	123	80.95	5181301	93.64	21775	74.52	92.37	57.94
Real Estate	261	24.22	5592248	20.77	2862	62.34	80.87	58.54
Agriculture	10	73.97	3261259	99.14	6120	98.84	76.16	60.95
Arts & Entertainment	26	77.47	2709960	89.72	9663	79.00	73.05	63.37
Manufacturing	1754	81.88	8247890	89.65	9930	68.86	88.37	68.53
Wholesale	106	76.73	3935273	76.18	9604	85.55	87.24	69.02
Health Care	57	91.02	4376716	98.40	23052	89.08	85.22	70.64
Waste Management	75	85.74	5202655	82.67	25133	92.70	86.21	71.09
Construction	56	86.45	2720294	97.05	7277	92.22	78.56	71.61
Oil & Gas	223	73.22	4263378	87.07	4095	59.81	74.28	71.88
Hotel and Food	67	84.40	7501438	91.37	40331	84.42	84.88	73.50
Utilities	90	75.99	11912421	89.34	7180	66.29	85.25	74.20
Transportation	130	56.14	8082138	71.75	18584	74.20	88.04	77.55
Finance	743	70.83	7238875	88.30	8271	64.78	89.51	78.10
Retail	151	78.49	17221834	80.65	62861	60.59	95.07	78.40
All	4394	73.46	8371894	81.33	12588	67.15	88.37	69.22

Table 3: Correlation between Uncertainty Proxies

The table reports the time-series correlation between $\widetilde{RRA}_t$, the aggregate measure of firms' raw relative attention to uncertainty from equation (2), and various proxies for uncertainty proposed by the prior literature. These alternative proxies for uncertainty include the Chicago Board Options Exchange's CBOE Volatility Index (VIX), the economic policy uncertainty measure of Baker et al. (2016) (EPU), and the financial (FIN), macroeconomic (MAC), and real (REA) uncertainty measures from Ludvigson et al. (2021). We compute the correlation between $\widetilde{RRA}_t$ and each of VIX and EPU using weekly data, as these measures are available at high frequency. In contrast, since the measures from Ludvigson et al. (2021) are only available at the monthly frequency, we compute the correlation between $\widetilde{RRA}_t$ and these measures at the monthly frequency. The sample period is from 2016 through 2022.

	$\widetilde{\text{RRA}}$	VIX	EPU	FIN	MAC	REA
$\widetilde{\text{RRA}}$	1.000					
VIX	0.415	1.000				
EPU	0.438	0.618	1.000			
FIN	0.577	0.755	0.635	1.000		
MAC	0.636	0.752	0.778	0.746	1.000	
REA	0.678	0.723	0.765	0.727	0.989	1.000

Table 4: Sensitivity of Uncertainty-Related Reading to Fluctuations in Common Uncertainty Proxies

This table presents the sensitivity of firms' uncertainty-related readings to fluctuations in aggregate uncertainty. Each week, we first sort firms into three (Panel A) or five (Panel B) portfolios based on their adjsuted relative attention scores ($\widetilde{ARA_t}$) from the previous week, as measured by equation (6). We, then, define the portfolio-level raw relative attention score to uncertainty using equation (2) aggregating across all firms assigned to each portfolio. Finally, we estimate a time-series regression in which the dependent variable is the portfolio's raw relative attention to uncertainty, and the independent variable is one of five common proxies for aggregate uncertainty described in Table 3. The table reports each portfolio's β from the aforementioned regression and the difference in sensitivity between low- and high-attention-sorted portfolios. The sample period spans from 2016 to 2022, and t-statistics, in brackets, are computed using Newey and West (1987) standard errors.

(a) Three *ARA* Portfolios

	VIX		EPU		Financial		Macro		Real	
Portfolio	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
Low ARA	0.0178	[3.88]	0.0160	[4.07]	0.0214	[3.49]	0.0261	[4.04]	0.0268	[4.37]
2	0.0271	[4.04]	0.0292	[5.00]	0.0374	[3.87]	0.0414	[4.01]	0.0440	[4.54]
High ARA	0.0294	[3.90]	0.0322	[4.81]	0.0424	[3.86]	0.0473	[3.88]	0.0510	[4.46]
High-Low	0.0115	[3.23]	0.0162	[5.29]	0.0210	[3.81]	0.0212	[3.33]	0.0242	[4.24]

(b) Five *ARA* Portfolios

	VIX		EPU		Financial		Macro		Real	
Portfolio	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
Low ARA	0.0136	[3.62]	0.0098	[3.44]	0.0161	[3.67]	0.0194	[4.41]	0.0190	[4.68]
2	0.0218	[3.88]	0.0219	[4.47]	0.0271	[3.49]	0.0327	[3.93]	0.0342	[4.34]
3	0.0273	[4.10]	0.0293	[5.06]	0.0375	[3.84]	0.0417	[4.02]	0.0443	[4.55]
4	0.0295	[3.91]	0.0320	[4.87]	0.0425	[3.82]	0.0458	[3.82]	0.0491	[4.36]
High ARA	0.0290	[3.89]	0.0321	[4.80]	0.0416	[3.87]	0.0477	[3.94]	0.0515	[4.53]
High-Low	0.0154	[3.30]	0.0223	[5.23]	0.0255	[3.45]	0.0283	[3.18]	0.0326	[4.03]

Table 5: Risk Mitigation, Compliance, and Firm Uncertainty

This table presents regression results examining the relationship between firm-level hedging intensity (Panel A) and regulatory intensity (Panel B) with firm uncertainty by estimating equation (8). In Panel A, the dependent variable measures firm-level hedging intensity that we define using the method's described in Campello et al. (2011). Specifically, we count the number of hedging-related words each firm mentions in its annual 10-K and scale this quantity by the total number of words in the firm's 10-K. To eliminate the influence of outliers, we transform this firm-level hedging intensity measure into an indicator variable that takes on a value of one if the firm has a hedging-intensity ratio above the cross-sectional median and zero otherwise. In Panel B, we measure a firm's regulatory intensity using the measure constructed by Kalmenovitz (2022). Regressions include combinations of industry, date, and industry-by-date fixed effects, as well as control variables from Leary and Roberts (2014) and Whited and Wu (2006), and uncertainty-related controls. Panel A (Panel B) uses data from June 2016 through July 2022 (December 2020). All standard errors are clustered by firm and time.

	(1)	(2)	(3)	(4)	(5)
Panel A: Hedging Activity					
$ARA_{i,t-1}$	0.3282***	0.3345***	0.3165***	0.2973***	0.2839***
	[4.00]	[4.47]	[4.14]	[3.62]	[3.17]
Observations	10,437	10,437	10,362	7,990	6,531
R^2	0.0229	0.1479	0.2040	0.3694	0.3799
Panel B: Compliance Activity					
$ARA_{i,t-1}$	11.1793***	9.4434***	7.6316***	1.3079	2.0319*
	[4.56]	[3.66]	[3.74]	[1.33]	[1.84]
Observations	23,812	23,812	23,696	20,131	16,782
R^2	0.0069	0.1345	0.4266	0.7415	0.7529
Date FE		+	+		
Industry FE			+		
Date $\times$ Industry FE				+	+
Controls				+	+
Unc. Controls					+

Table 6: **Implied Cost of Capital and Firm Uncertainty**
This table shows the results of estimating equation (10) to examine the relation between a firm's implied cost of capital, measured using the approach of Gebhardt et al. (2001), and the firm's relative attention to uncertainty (ARA), measured via equation (6). Regressions include combinations of industry, date, and industry-by-date fixed effects. Additionally, specifications (5) and (6) feature the set of controls from Leary and Roberts (2014) and Whited and Wu (2006), and uncertainty-related controls. The data underlying this regression spans from 2016 through 2022, and all standard errors are clustered by firm and time.

	(1)	(2)	(3)	(4)	(5)
$\text{ARA}_{i,t-1}$	0.0217***	0.0213***	0.0150***	0.0071**	0.0080***
	[8.02]	[7.88]	[6.97]	[2.56]	[2.83]
$\text{Size}_{i,t-1}$				0.0053***	0.0068***
				[3.85]	[4.92]
$\text{Leverage}_{i,t-1}$				0.0201***	0.0198***
				[12.26]	[11.64]
$\text{ROA}_{i,t-1}$				0.0060***	0.0094***
				[3.02]	[3.49]
$\text{Tangability}_{i,t-1}$				−0.0022*	−0.0029**
				[−1.72]	[−2.37]
Tobin $\text{Q}_{i,t-1}$				−0.0030*	−0.0042**
				[−1.88]	[−2.62]
Fin. $\text{Constraints}_{i,t-1}$				−0.0008*	−0.0010**
				[−1.79]	[−2.12]
VIX $\beta_{i,t-1}$					0.0001
					[0.23]
Political $\text{Risk}_{i,t-1}$					−0.0000
					[−0.13]
Non-political $\text{Risk}_{i,t-1}$					0.0004
					[1.06]
Date FE		+	+		
Industry FE			+		
Date × Industry FE				+	+
Observations	36,573	36,573	36,455	29,255	26,825
R^2	0.0103	0.0172	0.0785	0.2264	0.2469

Table 7: Real Outcomes and Firm Uncertainty

This table reports the results of estimating equation (9) to examine the relation between a firm's relative attention to uncertainty (measured by the uncertainty index) and asset growth (Panel A), sales growth (Panel B), and employment growth (Panel C). Asset growth, sales growth, and employment growth are defined in the Internet Appendix OA.1. The specifications include combinations of industry, date, and industry-by-date fixed effects. Additionally, specification (5) features the set of controls from Leary and Roberts (2014) and Whited and Wu (2006), and uncertainty-related controls. The data underlying this regression spans from 2016 through 2022, and all standard errors are clustered by firm and time.

	(1)	(2)	(3)	(4)	(5)
			Panel A: Asset Growth		
$\text{ARA}_{i,t-1}$	-0.0394^{***}	-0.0391^{***}	-0.0266^{***}	-0.0112^{**}	-0.0132^{**}
	[−3.34]	[−3.41]	[−3.16]	[−2.22]	[−2.49]
Observations	52,794	52,794	52,393	40,059	33,241
R^2	0.0019	0.0149	0.0208	0.1117	0.1212
			Panel B: Sales Growth		
$\text{ARA}_{i,t-1}$	-0.1987^{***}	-0.1952^{***}	-0.1390^{***}	-0.0632^{***}	-0.0531^{***}
	[−7.41]	[−7.34]	[−6.97]	[−3.70]	[−3.00]
Observations	48,078	48,078	47,830	39,616	33,039
R^2	0.0030	0.0127	0.0367	0.0942	0.1102
			Panel C: Employment Growth		
$\text{ARA}_{i,t-1}$	-0.1999^{***}	-0.1956^{***}	-0.1726^{***}	-0.0741^{**}	-0.0584^{**}
	[−4.09]	[−4.05]	[−3.43]	[−2.17]	[−2.67]
Observations	10,335	10,335	10,260	7,947	6,517
R^2	0.0074	0.0194	0.0287	0.1123	0.1243
Date FE		+	+		
Industry FE			+		
Date × Industry FE				+	+
Controls				+	+
Unc. Controls					+

Table 8: Firm-Level Uncertainty Portfolios: Characteristics
This table presents the characteristics of portfolios formed from sorts on firm relative attention to uncertainty (ARA), measured via equation (6). We sort the cross-section of firms into tercile portfolios in Panel A and quintile portfolios in Panel B and rebalance the portfolios at the end of each quarter. The table reports the value-weighted average of five prominent asset-pricing characteristics: the market beta, market capitalization, book-to-market ratio, gross profitability rate, and asset growth rate of the underlying firms. In addition to reporting the average characteristic of each portfolio, the table also reports the difference in characteristics between the high ARA portfolio and the low ARA portfolio. The underlying data spans from 2016 through 2022, and brackets report t-statistics, which are constructed using Newey and West (1987) standard errors.

(a) Three ARA Portfolios

	Beta	Market Cap	Book to Market	Gross Profit	Asset Growth
Low ARA	1.0186	3439	0.5110	0.1742	0.3022
2	1.0557	5356	0.5291	0.2425	0.2150
High ARA	1.0628	11978	0.5313	0.2604	0.1453
High-Low	0.0443	8538	0.0203	0.0862	-0.1568
t-stat	[1.48]	[13.22]	[2.32]	[3.49]	[-3.49]

(b) Five ARA Portfolios

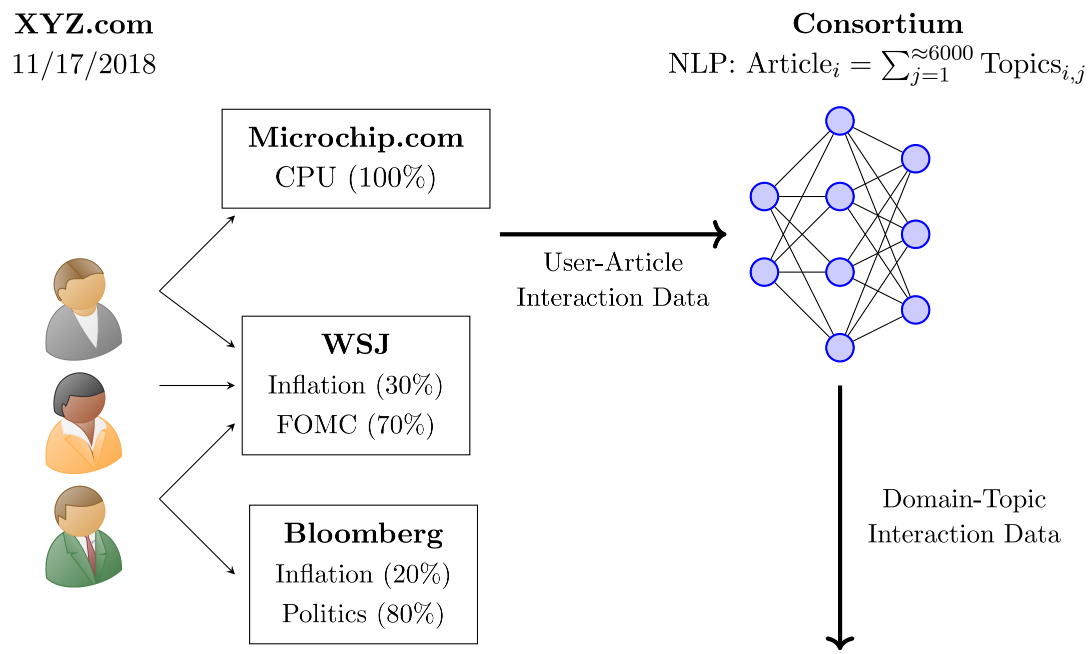
	Beta	Market Cap	Book to Market	Gross Profit	Asset Growth
Low ARA	1.0103	3157	0.5146	0.1588	0.3480
2	1.0336	3960	0.5052	0.2052	0.2380
3	1.0573	5448	0.5279	0.2434	0.2139
4	1.0747	7472	0.5419	0.2702	0.1575
High ARA	1.0525	14584	0.5293	0.2509	0.1472
High-Low	0.0422	11427	0.0147	0.0921	-0.2007
t-stat	[1.25]	[9.51]	[1.23]	[3.20]	[-3.64]

Table 9: Variance Decomposition by Fixed Effects and Firm Characteristics
This table presents a variance decomposition of the firm's relative attention to uncertainty (ARA) based on projections of ARA on fixed effects and firm characteristics. The variance of ARA is decomposed into variation coming from firm characteristics, sector fixed effects, sector-by-date fixed effects, and firm fixed effects. The firm characteristics we consider include market beta, size, book-to-market ratio, gross profitability, and asset growth, to the variance of ARA. The data spans from 2016 through 2022, with industry fixed effects estimated at two different levels of NAICS industry classification granularity.

	2-digit NAICS	3-digit NAICS	No Fixed Effect
Sector FE	3.78%	6.82%	
Sector $\times$ Date FE	1.32%	2.93%	
Firm-specific	94.89%	90.25%	
Permanent difference across firms, within sector-date	34.01%	30.88%	
Across firm-time residual	60.88%	59.37%	
Characteristics:			
Beta	0.22%	0.22%	0.19%
Size	0.01%	0.01%	8.91%
Book-to-Market	0.04%	0.04%	1.27%
Gross Profitability	0.10%	0.10%	0.29%
Asset Growth	0.01%	0.01%	1.13%
Characteristic Total	0.36%	0.38%	11.78%
Number of Sectors	19	72	

Figure 1: Visualizing the Consortium's Member Base

This figure illustrates a subsample of the Consortium's 4,000 members.



Date	Domain	Score	Topic
11/17/2018	xyz.com	110 $(3 \times 30 + 1 \times 20)$	Inflation
11/17/2018	xyz.com	210 (3×70)	FOMC
11/17/2018	xyz.com	100 (1×100)	CPU
11/17/2018	xyz.com	80 (1×80)	Politics

Figure 2: Visual Representation of the Consortium's data
The figure presents a stylized example of the data generation process within the Consortium's workflow. Initially, users read various articles from a range of online publishers. The online publishers then provide this user-interaction data to the Consortium. Utilizing advanced machine learning and natural language processing algorithms, the Consortium decomposes each article into a weighted-average combination of its core topics (illustrated beneath each article). Subsequently, the Consortium aggregates the analyzed user-article interaction data across users and firms (depicted as domains) to generate domain-topic interaction data. This data encompasses several variables, such as the date of interactions, the domain of interactions, the intensity of a domain's users engaging with a specific topic, and an associated topic label.

Figure 3: **Category Visualization**

This figure presents a word cloud illustrating the prevalence of topics across the Consortium's dataset, with each topic being associated with a category. Category labels are weighted by the number of associated topics, making categories with more topics appear larger. The visualization highlights the concentration of topics in categories such as finance, business services, and technology.

(a) Raw topic cloud

(b) ***tf-iaf*****-weighted topic cloud**

Figure 4: Visualizing the Uncertainty-Related Topics

This figure presents two topic clouds for the same set of uncertainty-related topics. Panel A displays the raw reading topic cloud, showing the most read topics in aggregate. Panel B presents the *tf-iaf*-weighted topic cloud, highlighting the topics that are more informative in distinguishing reading about uncertainty across firms in the cross-section. We define the corpus of uncertainty-related articles following the approach outlined in Section 2.2. Likewise, the *tf-iaf* weights are given by equation (5). The data underlying these topic clouds is from the week of 11/17/2018.

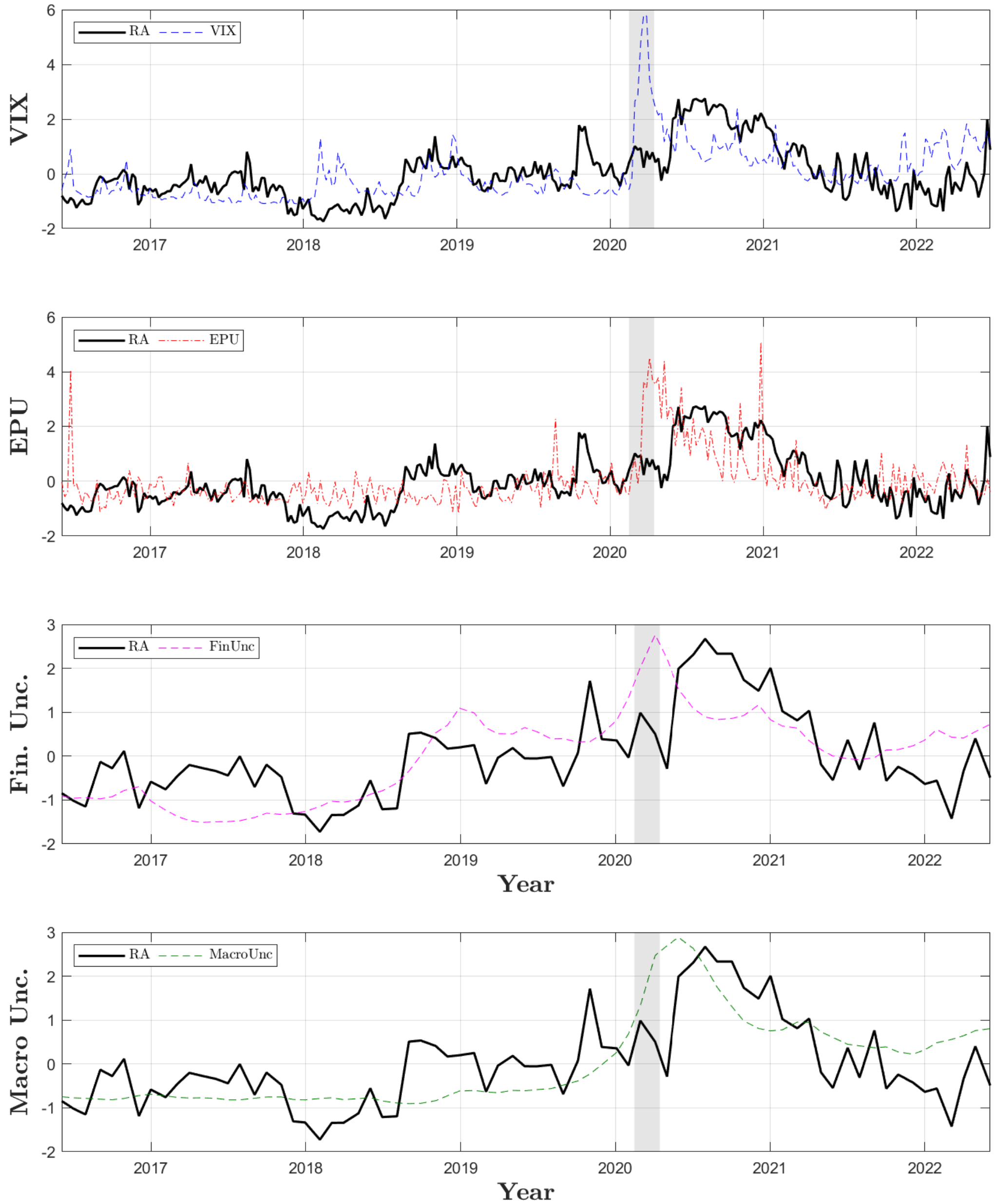
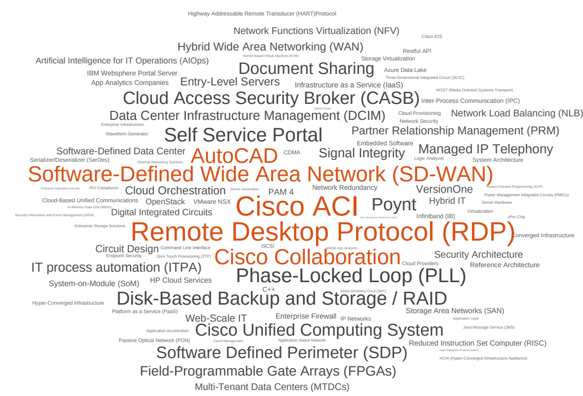


Figure 5: **Time Series of the Aggregate Relative Attention Measure $\widetilde{RRA}_t$**
This figure displays the time series of the aggregate measure of raw relative attention to uncertainty, $\widetilde{RRA}_t$, as defined in equation (2), alongside four established uncertainty proxies: the CBOE Volatility Index (VIX) in Panel A, the Economic Policy Uncertainty measure by Baker et al. (2016) (EPU) in Panel B, and the Financial and Macro Uncertainty measures from Ludvigson et al. (2021) in Panel C and D, respectively. The data spans from 2016 through 2022, with weekly time series for VIX and EPU and a monthly time series for the Ludvigson et al. measures.

(a) Raw topic cloud for Chemicals Manufacturing industry

(b) *tf-iaf*-weighted topic cloud for Chemicals Manufacturing industry

(c) Raw topic cloud for Computer and Electronics Manufacturing industry



(d) *tf-iaf*-weighted topic cloud for Computer and Electronics Manufacturing industry

Figure 6: Raw and *tf-iaf*-weighted topic clouds for two industries

This figure compares the raw and *tf-iaf*-weighted number of interactions across topics for the Chemicals Manufacturing (NAICS 325) and Computer and Electronics Manufacturing (NAICS 334) industries. The figure highlights the difference in topic relevance between the two weighting methods. The *tf-iaf* weights employed for each industry are given by equation (4). The data underlying these topic clouds is from the week of 11/17/2018.

A Online Appendix

OA.1 Variable Definitions

Asset growth. The asset growth rate is computed as the change in total assets (Compustat Quarterly item ATQ) between fiscal quarter t and $t-1$.

Asset tangibility. In analyses that feature quarterly data, asset tangibility is defined as the ratio of a firm's net property, plant, and equipment (Compustat Quarterly item PPENTQ) to the firm's total assets (Compustat Quarterly item ATQ) in fiscal quarter t. In analyses that feature annual data, asset tangibility is defined as the ratio of a firm's net property, plant, and equipment (Compustat Annual item PPENT) to the firm's total assets (Compustat Annual item AT) in fiscal year t.

Employment growth. The firm's employment growth rate is computed as the growth rate in the total number of employees (Compustat Annual item EMP) between fiscal years t and $t-1$.

Implied cost of capital (ICC). To construct this measure, we extract consensus estimates of earnings-per-share before extraordinary expenses ($FEPS_{t+i}$) at time $t+i$ from IBES, the book-value-per-share (B_t) at time t from Compustat, and the month-end stock price (P_t) at time t from CRSP. Our estimate of the ICC is then the solution to the following internal rate of return calculation:

$$P_t = B_t + \frac{FEPS_{t+1}/B_t - ICC_t}{(1+ICC_t)} \times B_t + \frac{FEPS_{t+2}/B_{t+1} - ICC_t}{(1+ICC_t)^2} \times B_{t+1} + TV_t. \tag{10}$$

The solution to this equation provides us with an estimate of a stock's ICC for each month, which is the frequency at which analysts update or reiterate their earnings forecasts in IBES.[17]

In general, only one- and two-year ahead earnings forecasts are reliably available. Hence, we project current earnings forward so as to estimate the terminal value term (i.e., TV_t) in equation (10). Following Pástor, Sinha, and Swaminathan (2008), we define the time-series of all analyst estimated implied return-on-equity (i.e., $FROE_{t+\tau} = FEPS_{t+\tau}/B_{t+\tau-1}$) such that they converge

[17] We rely on analyst forecasts rather than using predictive regressions (see, e.g. Hou, Van Dijk, and Zhang, 2012) due to our short sample and the large number of unanticipated events that occurred between 2016 and 2021 (e.g. the 2016 election results and the economic consequences of the COVID-19 pandemic).

to their long-run, pre-sample industry means (3-digit NAICS from 1995-2015) over a 15 year period,

$$TV_t = \sum_{\tau=3}^{14} \frac{FROE_{t+\tau} - ICC_t}{(1+ICC_t)^{\tau}} \times B_{t+\tau-1} + \frac{FROE_{t+15} - ICC_t}{ICC_t \times (1+ICC_t)^{14}} \times B_{t+14}, \tag{11}$$

where $FROE_{t+15}$ is the firm's industry (NAICS 3-digit) historical mean return-on-equity between 1995 and 2015 (i.e., before our sample). We follow the clean surplus methodology to estimate the future book values, where $B_{t+\tau} = B_{t+\tau-1} + FEPS_{t+\tau} \times (1 - \text{Payout Ratio})$. Like before, we estimate the historical long-run payout ratio using a firm's industry mean payout between 1995 and 2015.

Inventory growth. The inventory growth rate is computed as the change in inventories (Compustat Quarterly item INVTQ) between fiscal quarters t and $t-1$.

Investment growth. The physical investment growth rate is computed as the change in a firm's net property, plant, and equipment (Compustat Quarterly item PPENTQ) between fiscal quarters t and $t-1$.

Leverage. In analyses that feature quarterly data, leverage is defined as the sum of a firm's debt in current liabilities (Compustat Quarterly item DLCQ) plus long-term debt (Compustat Quarterly item DLTTQ) scaled by the firm's total assets (Compustat Quarterly item ATQ) in fiscal quarter t. In analyses that feature annual data, leverage is defined as the sum of a firm's debt in current liabilities (Compustat Annual item DLC) plus long-term debt (Compustat Annual item DLTT) scaled by the firm's total assets (Compustat Annual item AT) in fiscal year t.

Non-political risk. We measure each firm's non-political risk via the method developed by Hassan et al. (2019), who apply tools from computational linguistics to analyze the earnings call transcripts of each firm. This data is made available from `https://www.firmlevelrisk.com/`.

Political risk. We measure each firm's political risk via the method developed by Hassan et al. (2019), who apply tools from computational linguistics to analyze the earnings call transcripts of each firm. This data is made available from `https://www.firmlevelrisk.com/`.

Profitability. In analyses that feature quarterly data, profitability in fiscal quarter t is defined as the ratio of net income in fiscal quarter t (Compustat Quarterly item NIQ) to total assets in fiscal quarter $t-1$ (Compustat Quarterly item ATQ). In analyses that feature annual data, profitability in fiscal year t is defined as the ratio of net income in fiscal year t (Compustat Annual item NI) to total assets in fiscal year $t-1$ (Compustat Annual item AT).

Size. In analyses that feature quarterly data, firm size is measured using the natural logarithm of total sales (Compustat Quarterly item SALEQ) in fiscal quarter t. In analyses that feature annual data, firm size is measured using the natural logarithm of total sales (Compustat Annual item SALE) in fiscal year t.

Tobin's Q. In analyses that feature quarterly data, Tobin's q in fiscal quarter t is defined as the sum of the book value of total assets (Compustat Quarterly item ATQ) minus the book value of common equity (Compustat Quarterly item CEQQ) plus the market value of common equity (Compustat Quarterly item CSHOQ multiplied by the end of fiscal quarter stock price given by Compustat Quarterly item PRCCQ), all divided by the book value of total assets. In analyses that feature annual data, Tobin's q in fiscal year t is defined as the sum of the book value of total assets (Compustat Annual item AT) minus the book value of common equity (Compustat Annual item CEQ) plus the market value of common equity (Compustat Annual item CSHO multiplied by the end of fiscal year stock price given by Compustat Annual item PRCC), all divided by the book value of total assets.

VIX $\boldsymbol{\beta}$. We estimate each firm's exposure to fluctuations in the VIX index via the following regression:

$$R^e_{i,t} = \beta_0 + \beta_1 \text{MKTRF}_t + \beta_2 \Delta\text{VIX}_t + \varepsilon_{i,t}.$$

Here, $R^e_{i,t}$ refers to the daily stock return of firm i on day t, MKTRF_t is the daily excess stock return of the aggregate market portfolio, measured using the CRSP value-weighted index, and ΔVIX_t is the difference in the level of the VIX index between days $t-1$ and t. In estimating this regression we a quarter's worth of daily data that precedes date t.

OA.2 Additional Tables and Figures

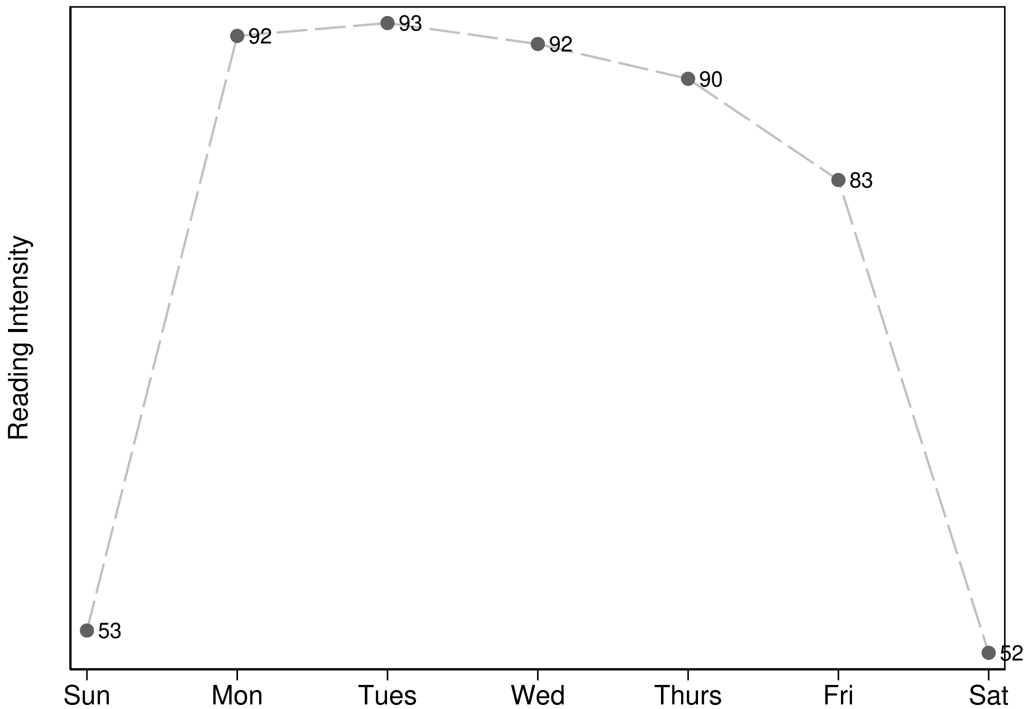


Figure OA.2.1: Average number users per day of the week
The figure plots the average number of unique users at each firm that interact with the Consortium's data on each day of the week. The set of domains over which these numbers are averaged cover both private and public firms from 2016 to 2022. To address the intra-week patterns in reading, the daily domain-topic interaction data is aggregated into lower frequencies, such as weekly, monthly, or quarterly, for analysis in the main paper.

Table OA.2.1: Summary Statistics (Controls).
The table presents the summary statistics associated with the key real outcomes and control variables employed in Section 4. Here, "N" refers to the total number of observations for each variable, "SD" denotes the variable's standard deviation, and "p25" ("p75") refers to the 25^{th} (75^{th}) percentile of the variable's distribution. All data are quarterly, except for the employment growth rate, which is annual, and spans 2016 through 2022.

	N	Mean	SD	p25	Median	p75
Asset growth	52764	0.0363	0.1985	-0.0315	0.0042	0.0405
Employment growth	11554	0.0680	0.2949	-0.0476	0.0215	0.1224
Sales growth	48320	0.0882	0.4823	-0.0576	0.0244	0.1153
Log(Sales)	48640	4.7980	2.6002	3.3425	5.1538	6.5673
Tobin's q	52764	2.5606	2.2098	1.2267	1.7824	3.0016
Profitability	49521	-0.0296	0.1021	-0.0376	0.0032	0.0188
Leverage	52454	0.2695	0.2432	0.0529	0.2353	0.4084
Tangibility	51965	0.2378	0.2417	0.0605	0.1437	0.3355

Table OA.2.2: Raw Reading and Fluctuations in Common Uncertainty Proxies: A Placebo Test

This table displays the sensitivity of firms' raw uncertainty-related reading to fluctuations in aggregate uncertainty. In each week of the sample period we sort the cross-section of firms into either three (Panel A) or five (Panel B) portfolios based on their relative raw attention (RRA) scores from the previous week. We then compute the portfolio's relative attention to uncertainty via equation (2), where we compute the summation across all firms assigned to a specific portfolio. Finally, we compute the time-series regression (7) in which the dependent variable is the portfolio's relative attention to uncertainty and the independent variable is one of five common proxies for aggregate uncertainty: the Chicago Board Options Exchange's CBOE Volatility Index ("VIX"), the economic policy uncertainty measure of Baker et al. (2016) ("EPU"), and the financial, macroeconomic, and real uncertainty measures from Ludvigson et al. (2021) that we denote by "Financial," "Macro," and "Real," respectively. The table reports each portfolio's β from this regression, as well as beta of the difference in the sensitivity of reading across the low and high attention-sorted portfolios. The time series of data underlying each regression spans from 2016 through 2022 and brackets report t-statistics that are computed using Newey and West (1987) standard errors.

(a) Three RRA Portfolios

	VIX		EPU		Financial		Macro		Real	
Portfolio	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
Low ARA	0.0355	[5.14]	0.0397	[5.13]	0.0443	[4.19]	0.0378	[2.36]	0.0422	[2.80]
2	0.0359	[4.89]	0.0401	[4.72]	0.0473	[3.59]	0.0461	[2.66]	0.0508	[3.09]
High ARA	0.0321	[4.64]	0.0366	[4.98]	0.0507	[4.00]	0.0483	[3.55]	0.0510	[3.90]
High-Low	-0.0034	[-0.62]	-0.0031	[-0.62]	0.0064	[0.61]	0.0105	[1.01]	0.0088	[0.86]

(b) Five RRA Portfolios

	VIX		EPU		Financial		Macro		Real	
Portfolio	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
Low ARA	0.0315	[4.78]	0.0354	[4.91]	0.0371	[3.84]	0.0305	[2.07]	0.0345	[2.49]
2	0.0406	[5.36]	0.0452	[5.26]	0.0539	[4.36]	0.0484	[2.67]	0.0534	[3.13]
3	0.0356	[4.82]	0.0398	[4.68]	0.0468	[3.52]	0.0457	[2.62]	0.0504	[3.04]
4	0.0339	[4.67]	0.0370	[4.47]	0.0454	[3.28]	0.0469	[2.92]	0.0508	[3.30]
High ARA	0.0309	[4.53]	0.0363	[5.26]	0.0539	[4.44]	0.0489	[3.97]	0.0508	[4.29]
High-Low	-0.0007	[-0.11]	0.0009	[0.16]	0.0167	[1.41]	0.0184	[1.63]	0.0163	[1.48]

Table OA.2.3: Real Outcomes and Firm Uncertainty: Additional Evidence
The table reports the results of estimating equation (9) to examine the relation between a firm's relative attention to uncertainty and inventory growth (Panel A) and the growth rate of net property, plant, and equipment (Panel B). Regressions include combinations of industry, date, and industry-by-date fixed effects. Moreover, certain regressions feature the set of controls from Leary and Roberts (2014) and Whited and Wu (2006), and uncertainty-related controls described in the main text. The data underlying this regression spans from 2016 through 2022 and all standard errors are clustered by firm and time.

	(1)	(2)	(3)	(4)	(5)
Panel A: Inventory growth					
$\text{ARA}_{i,t-1}$	-0.0457^{***}	-0.0453^{***}	-0.0394^{***}	-0.0176^{**}	-0.0198^{***}
	[−6.16]	[−6.18]	[−5.60]	[−2.65]	[−2.87]
Observations	36,841	36,841	36,718	30,584	25,873
R^2	0.0025	0.0161	0.0235	0.1032	0.1191
Panel B: PPENT Growth					
$\text{ARA}_{i,t-1}$	-0.0979^{***}	-0.0982^{***}	-0.0692^{***}	-0.0368^{*}	-0.0458^{*}
	[−2.72]	[−2.94]	[−3.14]	[−1.86]	[−1.89]
Observations	51,976	51,976	51,634	39,760	33,116
R^2	0.0029	0.0877	0.0948	0.1987	0.2315
Date FE		+	+		
Industry FE			+		
Date × Industry FE				+	+
Controls				+	+
Unc. Controls					+

OA.3 Detailed Description of the Raw Attention Data

In this section we provide a detailed overview of the Consortium's *domain-topic* dataset, which is used in the bulk of our empirical analyses. The main takeaway from this section is that while the Consortium's dataset covers a wide variety of topics (more than 7,000 in 2022), the majority of these topics are generally uninformative about each firm's business line(s). With this in mind, Section 2.2 appeals to the growing literature in machine learning and computer science to propose a data-driven method for dealing with these uninformative, general, topics that encompass topics related to politics, entertainment, and sports.

To illustrate the fact that the vast majority of domain-topic interactions are only marginally informative about what a firm does, Figure OA.3.2 explores the distribution of topics for the Computer and Electronics Manufacturing sector (3-digit North American Industry Classification System (NAICS) code 334) during the week ending on November 17, 2018, a week that falls roughly in the midpoint of our sample period. The y-axis in this figure is a normalized measure of the intensity with which the employees of the firms in this sector are interacting with each topic in the given week. This normalized measure of the attention allocated to topic t is defined as

$$\text{NormInteractions}_t = \frac{\sum_{i\in\mathcal{I}} \text{Interactions}_{i,t}}{\max_t \left(\sum_{i\in\mathcal{I}} \text{Interactions}_{i,t}\right)}, \tag{12}$$

where $\text{Interactions}_{i,t}$ corresponds to the number of unique users at each firm i interacting with topic t in the given week. As this measure is scaled by the maximum number of unique interactions across all topics in a given week, $\text{NormInteractions}_{i,t}$ is a scalar that ranges from one (for the topic with the most interactions in a given week) to zero (for any topics with zero interactions in the given week). Numbers between these two extremes represent the amount of attention a given topic receives *relative* to the topic with the highest number of interactions in that week. The x-axis in this figure is then the rank associated with each topic's normalized attention score. These ranks are ordered from the topic with the most interactions, which has a rank of one, to the topic with the least number of interactions. In presenting these normalized interactions we truncate the rank at 250 to highlight the steep decline in attention as we move from the most popular topic to the less popular topics in a given week.

The topics with the most interactions towards the left of the figure are related a group of topics we consider to be current events. These high-interaction topics, which include "South by Southwest," "Call of Duty," and "US Secret Service," were highly relevant topics in the week underlying this exercise and appeared, for example, as headlines on the splash page of major publishers, such as USA Today. Figure OA.3.3 shows an example of three article headlines published by the Consortium's members for three of the top 10 topics underlying Figure OA.3.2.

The first headline highlights a pitch submission deadline for an event that takes place in March 2019 at the South By Southwest music festival, an event with a heavy tech presence that attracts hundreds of thousands of attendees each year. The second headline is regarding a new multiplayer map in the newest version of Call of Duty, a popular online video game. Finally, during the week in question, former President Trump was in Europe. A controversy erupted when the US Secret Service suggested that then President Trump avoid an event due to inclement weather. This goes to show that many of the topics with a high number of interactions in any given week very likely reflect news and current affairs. Untabulated analyses result in similar takeaways when we focus on sectors other than technology and different points in time.

As we move to the right of Figure OA.3.2, and the rank increases along the x-axis, we see a steep decline in the relative amount of attention paid to the topics with a rank between 150 to 250. Although these topics still draw more user interactions than the 5,750 or so other topics in the Consortium's dataset in the given week that we do not visualize, these topics with a rank between 150 and 250 still only attract about 10% of the interactions dedicated to the more popular current events described above. Yet, these topics are still very general in nature and cover "Miami, Florida," a popular retirement destination, "Traditional IRA," a common retirement savings account, and "Environment for Aging," an event on senior living design. These results once again highlights how many of the topics with a high number of interactions are very general in nature and do not necessarily reflect details on the business line(s) of the underlying firms.

This begs the question, can we use the Consortium's data to glean any novel insights about a firm's attention to economically relevant news, such as uncertainty, when most employees' attention is concentrated on common and current events? To demonstrate that the answer to this question is "yes," Figure OA.3.4 presents a histogram of interactions across the entire distribution of topics for firms in the computer and electronics sector (NAICS 334). Here, the x-axis reports the degree of topic interaction from topics with the least (zero) to most (one) interactions. The y-axis displays the proportion of topics that fall within some topic interaction interval.

Figure OA.3.4 indicates that the vast majority of the mass of the topic distribution is concentrated among topics that have a relatively low number of interactions. In fact, about 85% of the topics that firms in the computer and electronics sector interacted with during the week of 11/17/2018 received less than than 20% of the interactions dedicated to the 10 most popular topic in the sector that week (i.e., "Live Streaming," "Google +," "South By Southwest," "Call of Duty," " US Secret Service," etc.). For instance, topics such as "disk-based backup and storage," "circuit design" and "cloud access security broker" that are inherently related to firms in the computer and electronics sector each received about 5% or less of the interactions dedicated to "South by Southwest" and "Call of Duty" that week. Likewise, while topics related to business-relevant risks that we want to focus on, such as "credit risk," "exchange rate," and "cost of capital," received relatively more attention than "circuit design," these uncertainty-related topics still received far less attention than many current events.

Overall, Figure OA.3.2 confirms that the Consortium's data can indeed help us to glean novel insights about a firm's attention to economic uncertainty, provided that we are careful to account for the fact that the bulk of the average employee's attention is, unsurprisingly, dedicated to reading about current events. A corollary from this figure is that simple metrics of firm attention, such as the amount of total reading per employee, are very likely uninformative about the economic environment in which the firm is operating.

Consequently, Section 2.2 develops an intuitive measure of a firm's attention to uncertainty that is immune to the aforementioned concern in two steps. First, we define a set of topics that reflect uncertainty-related news, articles, and events. Second, we develop a measure of attention to these uncertainty-related topics that implicitly down weights interactions with topics that are common across all firms. This weighting scheme is motivated by the large literature on natural language processing (e.g., Gentzkow et al. (2019)) and essentially down weights a firm's attention to topics that all other firms are also reading about (e.g., "Call of Duty") and up weights topics that are more likely economically relevant and firm-specific (e.g., "Credit Risk" for all firms and "Circuit Design" for firms in the computer and electronics sector).

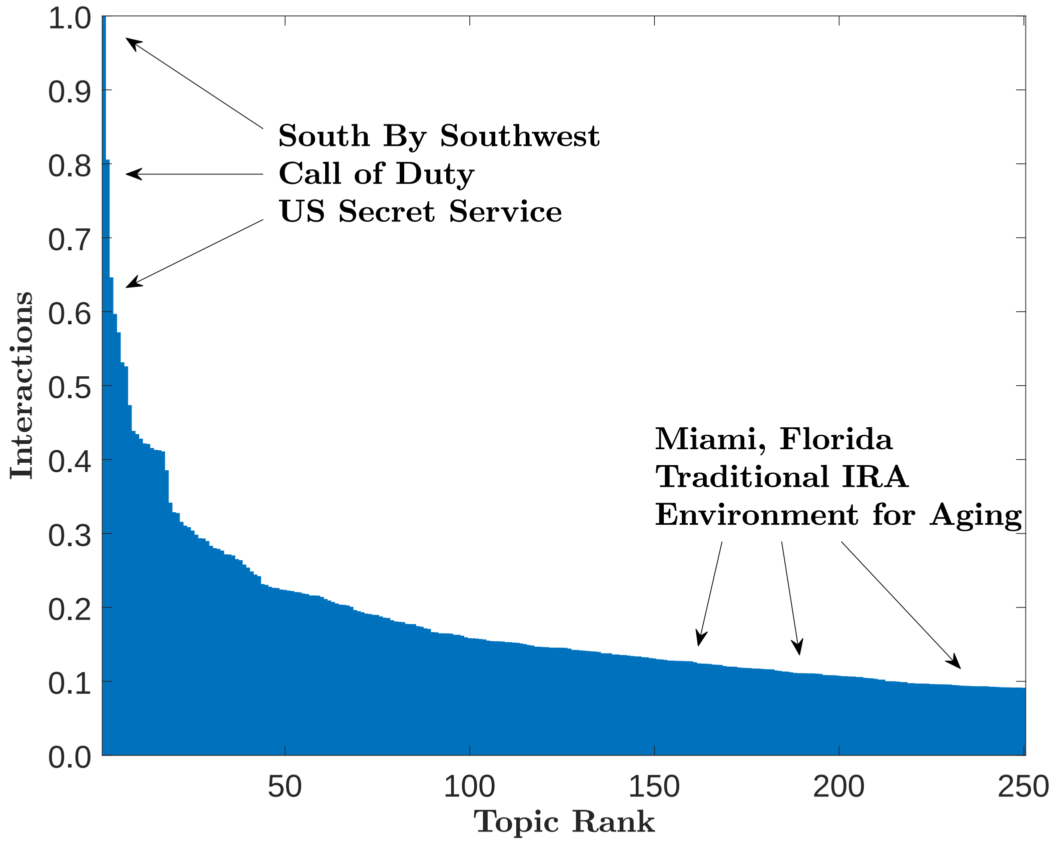


Figure OA.3.2: Bar chart of normalized attention to topics.
This bar chart represents the normalized number of interactions, defined according to equation (12), on the y-axis and rank of various topics across all firms in the North American Industry Classification (NAICS) code 334 industry — the Computer and Electronics Manufacturing industry — during the week ending 11/17/2018 on the x-axis.

(a) *South By Southwest Pitch Event*

Startup Funding Options, Media Exposure & More at SXSW Pitch – Final Deadline November 18

By Jordan Roberts 11/8/2018

(b) *Call of Duty Release*

NOVEMBER 13, 2018

by Call of Duty Staff

NUKETOWN COMES TO BLACK OPS 4

Forecast for November 20: Snowy, with a chance of mushroom clouds. Bundle up and hunker down, because Nuketown is coming back and it's better than ever.

(c) *US Secret Service News*

President Trump blames Secret Service for canceling cemetery trip in France

David Jackson

USA TODAY

Published 9:02 a.m. ET Nov. 13, 2018 | **Updated 1:46 a.m. ET Nov. 14, 2018**

Figure OA.3.3: News Headlines from 11/08/2018.
This figure captures a sample of headlines from Consortium publishers associated with 3 top topics from the week ending 11/08/2018. Panel (a) is a headline highlighting a submission deadline for a competitive South By Southwest VC pitch competition. Panel (b) highlights the release of an addition to the new release of Call of Duty. Panel (c) is a current events article about the US Secret Service.

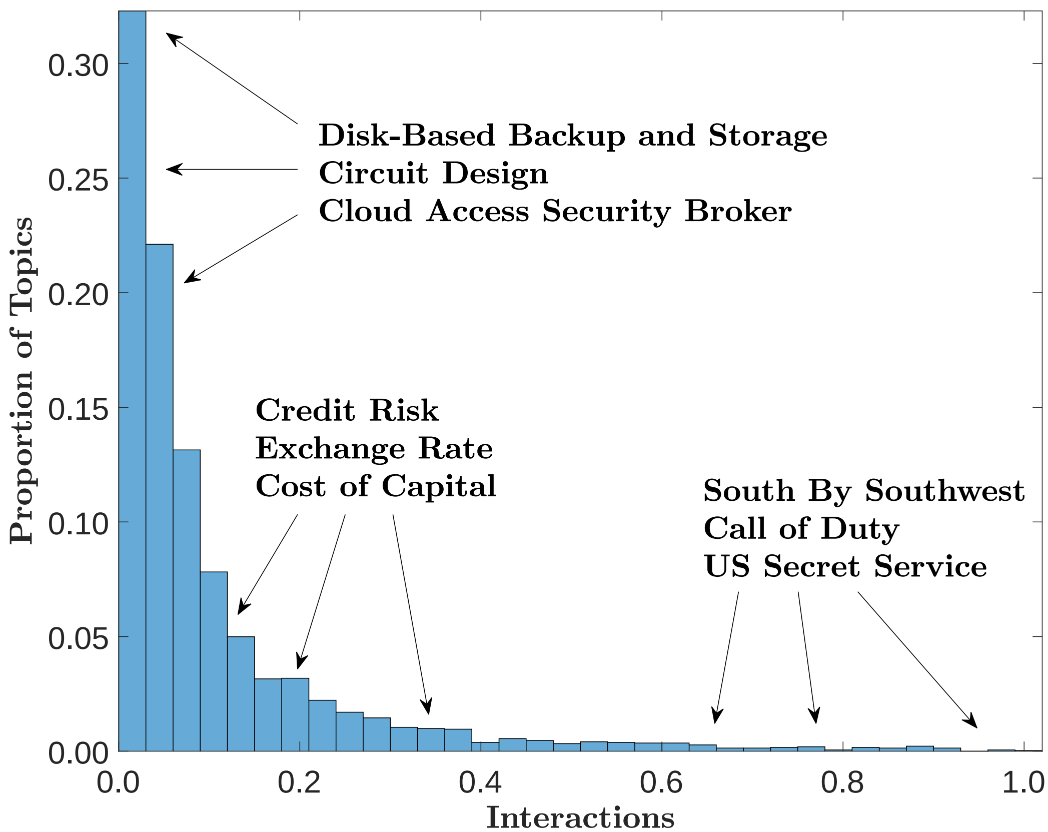


Figure OA.3.4: **Histogram.**
This figure is the histogram of topic interactions for all firms in NAICS 334 (Computer and Electronics Manufacturing) during the week ending 11/17/2018. We normalize the number of interactions with each topics by the number of interactions with the most popular topics in the given week (see equation (12). This measure of normalized interactions ranges from zero, for any topics with no interactions, to one, for the single topic with the most interactions.

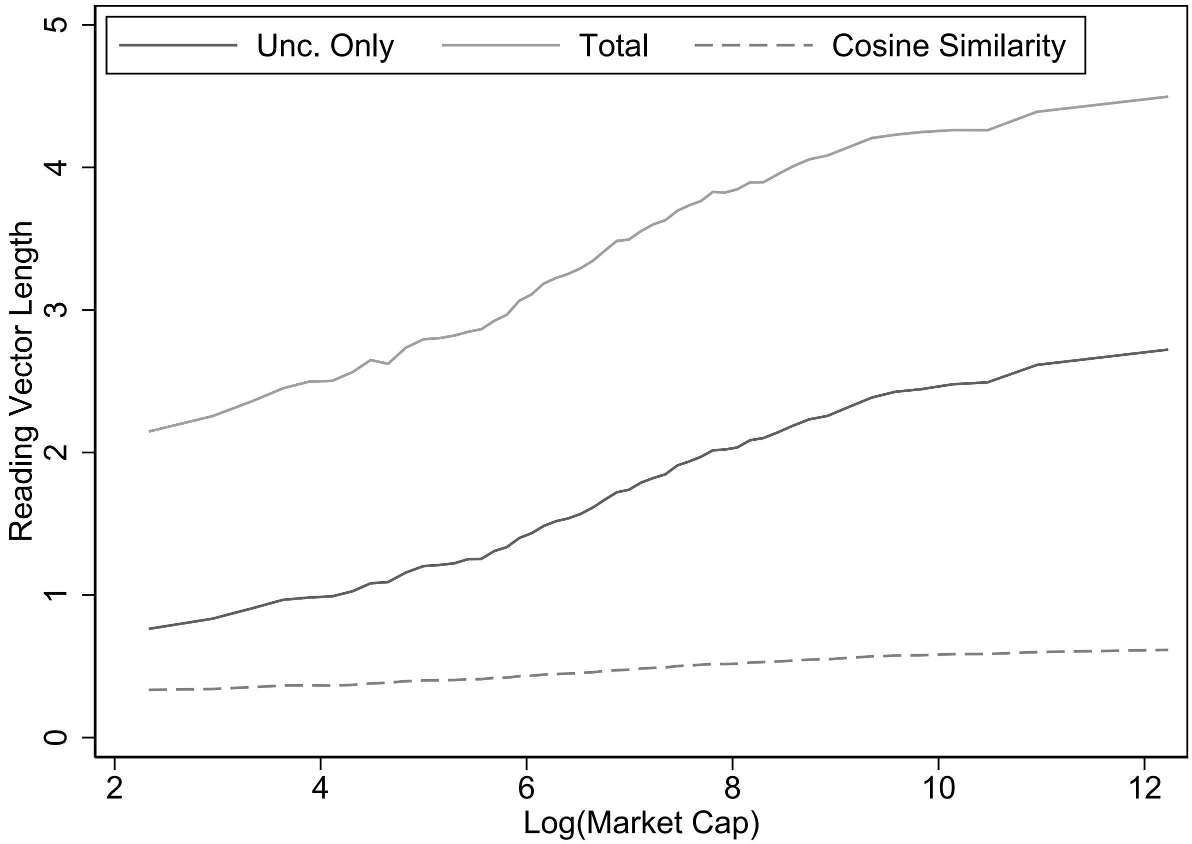


Figure OA.3.5: Relationship between Firm Size and Reading.
The figure presents the relation between firm reading and firm size, measured by the natural logarithm of a firm's market capitalization. We measure the amount of reading a firm allocates towards uncertainty-related and total news. Specifically, we compute the amount of reading by computing the average length of $\boldsymbol{tf}_{i,t}^{Unc}$ and $\boldsymbol{tf}_{i,t}^{Total}$ from equation (1). We then plot the length of each of these vectors as a function of firm size. Beyond plotting the amount of reading firms of various sizes allocate to each type of reading, the figure also plots the cosine similarity between the two vectors as a function of firm size. The data underlying this analysis ranges from 2016 through 2022.

OA.4 Stylized Framework

In this section we motivate why would one expect attention to uncertainty to be closely tied to a firm's expected rate of return. We begin with the canonical present value of equity expression:

$$P_t = E_t \left[\sum_{\tau=1}^{\infty} \frac{D_{t+\tau}}{\prod_{s=1}^{\tau} R_{t+s}} \right],$$

where P_t is the share price of a firm at time t, $D_{t+\tau}$ is the cash flow to equity holders at time $t+\tau$, and R_{t+s} is the gross discount rate at time $t+s$. To simplify the intuition, we log linearize the expression using the Campbell and Shiller (1988) approximation,

$$\begin{aligned} p_t &= d_t + E_t \left[\sum_{\tau=1}^{\infty} \rho^{\tau-1} \Delta d_{t+\tau} \right] - E_t \left[\sum_{\tau=1}^{\infty} \rho^{\tau-1} r_{t+\tau} \right] \\ &= (1-\rho) E_t \left[\sum_{\tau=1}^{\infty} \rho^{\tau-1} d_{t+\tau} \right] - E_t \left[\sum_{\tau=1}^{\infty} \rho^{\tau-1} r_{t+\tau} \right]. \end{aligned}$$

Here, $r_{t+\tau}$ is the log discount rate, $d_{t+\tau}$ is the log dividend, p_t is the log price, $\Delta d_t = d_t - d_{t-1}$, and ρ is the price-dividend ratio about which the approximation is linearized. As in Cochrane (2008) we ignore the mean of the approximation for simplicity.

We assume that a firm can optimally change r_t as a function of its hedging activity. Specifically, a firm can reduce its exposure to systematic risk through actions taken by its risk management group.[18] As highlighted by Brown (2001) and others, hedging can be costly; not only does it require substantial physical and intellectual capital to manage a portfolio of, e.g., derivatives, but the quantity of risk for a firm is also inherently unknowable. For example, a farmer today can easily hedge the *price* they will receive for a bushel of corn tomorrow, but is unable to perfectly hedge the *quantity* of corn they will produce due to uncertainty in weather, blight or other risks. Additionally, due to mark-to-market accounting for many hedging instruments, there is uncertainty in future margin requirements (i.e., liquidity risk).[19] The resulting trade-off limits the degree to which a firm will fully hedge its expected exposure, a notion that is critical for our empirical strategy. Going forward, rather than explicitly modeling the source of these cost we embed its certainty equivalence as a negative cash flow that reduces d_t (see Brown, 2001).

[18] A large literature discusses the incentives to hedge systematic exposure and its impact on firm cost of capital (see, inter alia, Bartram, Brown, and Conrad, 2011; Stulz, 1984).

[19] A recent Washington Post article highlights how uncertain margin requirements can affect firm prospects.

To glean the extent to which a given firm is trying to hedge the effects of uncertainty on its cost of capital, we measure the degree to which the firm is paying attention to uncertainty-related news relative to other important business activities. As we discuss in the next section, the data we use to study this question is obtained from a proprietary data vendor that tracks firm-level attention to a broad set of general products and services that other organizations sell. As both d_t and r_t are decreasing functions of a firm's hedging activities one can write them as function of the firm's relative attention to uncertainty versus other business-related topics (i.e., RA_t).

We define $\bar{r}_t$ as the discount rate prior to any hedging. This quantity is a function of the operating and financial leverage of the firm and is thus referred to as the firm's fundamental discount rate. We define a hedging function h_t that intuitively links higher relative attention to uncertainty to increased hedging of the firm's discount rate through the following parameterization:

$$r_t(\mathrm{RA}) = (1 - h_t(\mathrm{RA}))\bar{r}_t. \tag{13}$$

Thus, when hedging activity increases, the observable discount rate decreases away from $\bar{r}_t$. $h_t(\mathrm{RA})$ is a smooth, increasing function that is bounded between 0 and 1 (e.g., a logistic function). If we differentiate the price of the firm with respect to RA_t we obtain

$$\frac{\partial p_t}{\partial \mathrm{RA}_t} = \underbrace{(1-\rho)\sum_{\tau=1}^{\infty}\rho^{\tau-1}\frac{\partial d_{t+\tau}}{\partial \mathrm{RA}_t}}_{-C'(\mathrm{RA}_t)} - \underbrace{\sum_{\tau=1}^{\infty}\rho^{\tau-1}\frac{\partial r_{t+\tau}}{\partial \mathrm{RA}_t}}_{-B'(\mathrm{RA}_t)},$$

where $C'(\mathrm{RA}_t)$ and $B'(\mathrm{RA}_t)$ represent the marginal costs and benefits of hedging to the firm. As $\partial r_t/\partial \mathrm{RA}_t \leq 0$ reading about uncertainty translates into hedging activity that reduces the discount rate. This hedging activity, however, is costly and naturally reduces cash flows, i.e., $\partial d_t/\partial \mathrm{RA}_t \leq 0$.

The firm's problem is to choose how much relative attention to pay to uncertainty (e.g., the number of employees dedicate towards hedging activity) such that firm value is maximized. The firm therefore sets the marginal cost equal to the marginal benefit:

$$C'(\mathrm{RA}_t^*) = B'(\mathrm{RA}_t^*) = \sum_{\tau=1}^{\infty}\rho^{\tau-1}\bar{r}_\tau h_\tau'(\mathrm{RA}^*). \tag{14}$$

We further assume that $C(\mathrm{RA}_t)$ is convex and that $C(\mathrm{RA}_t)$ and $B(\mathrm{RA}_t)$ have a single crossing such that there is an internal optimal level of relative attention (RA_t^*). Figure OA.4.6 captures the simple intuition of this result. The x-axis represents possible relative attention choices. The firm will optimally allocate attention within the feasible region according to equation (14).[20] One important characteristic of $B(\mathrm{RA}_t)$ highlighted in the figure is that it must be more convex initially than $C(\mathrm{RA}_t)$. To satisfy single crossing, however, at optimal RA_t^* $B(\mathrm{RA}_t)$ will be concave.

A set of implicit derivatives provides us with intuition about this equilibrium. First, we take the derivative of equation (14) with respect to RA_t^* to obtain,

$$\frac{\partial \mathrm{RA}^*}{\partial \bar{r}_t} = \frac{\rho^{t-1} h_t'(\mathrm{RA}^*)}{C''(\mathrm{RA}^*) - B''(\mathrm{RA}^*)} \geq 0 \quad , \tag{15}$$

where $B''(\mathrm{RA}^*)$ is concave at the optimal amount of relative attention due to the single crossing assumption. Intuitively, this relationship arises due to higher discount rates providing stronger hedging incentives for the firm. Second, we take the derivative of equation (13) with respect to the fundamental discount rate,

$$\frac{\partial r_t}{\partial \bar{r}_t} = (1 - h(\mathrm{RA}_t)) - \bar{r}_t h'(\mathrm{RA}_t^*) \frac{\partial \mathrm{RA}_t^*}{\partial \bar{r}_t} \geq 0. \tag{16}$$

Intuitively, this relationship is the result of the trade off described earlier–i.e., that firms are unable to hedge away all uncertainty due to prohibitive costs.[21] This relationship implies that firms with higher fundamental discount rates should also have higher observable discount rates.

One key issue is that we cannot directly observe the fundamental discount rate. As we move to our empirical analysis, which is conducted in the cross-section of stocks, we need an additional requirement–that the hedging and the cost functions are similar across firms. Armed with this assumption, consider two firms, denoted by the superscripts i and j, with different fundamental discount rates $\bar{r}_t^i > \bar{r}_t^j$. First, from equation (15), the firm with the higher fundamental discount rate optimally pays relatively more attention to uncertainty:

$$\mathrm{RA}^{*i} > \mathrm{RA}^{*j}. \tag{17}$$

[20] In order to avoid corner solutions it must be that (i) $C(\mathrm{RA}_t)$ has to be convex enough otherwise the firm can mitigate all risk, (ii) $C(\mathrm{RA}_t)$ must also not be too convex otherwise there will be no incentive to hedge, and (iii) $B(\mathrm{RA}_t)$ at optimal relative attention must not be too concave otherwise the firm will mitigate all risk at zero cost.

[21] Equation (16) is satisfied if an "ongoing concern" condition holds:

$$1 - h(\mathrm{RA}_t^*) > \frac{h'(\mathrm{RA}_t^*) \times \bar{r}_t}{C''(\mathrm{RA}^*) - B''(\mathrm{RA}^*)}.$$

Assuming the the firm is infinitely lived the denominator on the RHS of the inequality, which is an infinite sum of variables that are greater than zero should drive the RHS to near zero.

Second, from equation (16), the firm with the higher fundamental discount rate also has the higher optimal relative attention:

$$r_t^i > r_t^j. \tag{18}$$

The model thus predicts that higher observable discount rates and relative attention, which is observable through our data, are positively related in the cross-section of stocks. After introducing our novel data in the next section we test this prediction and its implications for real outcomes in Sections 5 and 4.

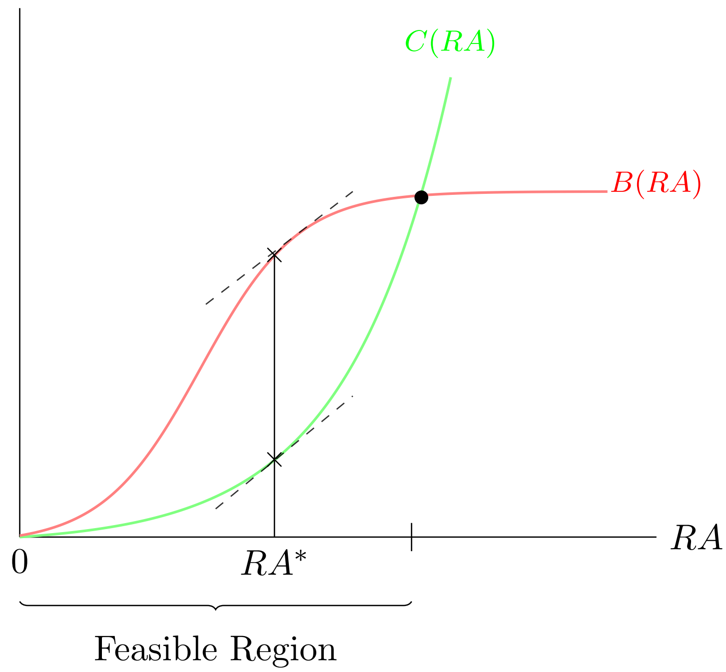


Figure OA.4.6: Stylized Example of Optimal Hedging
This figure presents a stylized general parameterization of the benefit and cost functions, highlighting key features of our theoretical framework.

Table OA.4.4: Firm-Level Uncertainty Portfolio: Risk Exposures and Pricing Errors
This table presents the risk exposures and pricing errors of a long-minus-short portfolio formed on the basis of each firm's relative attention to uncertainty, measured using equation (6). Specifically, we sort the cross section of firms at the end of each quarter t into 5 portfolios formed on the basis of $ARA_{i,t}$. We hold each portfolio through the following quarter end, at which point in time all portfolios are rebalanced. We compute the value-weighted return of a self-financing portfolio that buys the set of firms with the highest quintile $ARA_{i,t}$ and sells the set of firms with the lowest quintile $ARA_{i,t}$. We then regress the value-weighted returns of these portfolios onto a constant (Column (1)), the excess market return (Column (2)), the Fama and French (1993) three-factor model (Column (3)), Fama and French (2015) five-factor model (Column (4)), and the Fama and French (2015) five-factor model that also features the momentum factor (Column (6)). The table reports exposure of the portfolio to each risk factor as well as the portfolio's pricing error (α). The time-series regression is estimated using weekly data from 2016 through 2022 and t-statistics, reported in brackets, are computed using Newey and West (1987) standard errors.

	(1)	(2)	(3)	(4)	(5)
α	0.0014	0.0018	0.0016**	0.0006	0.0006
	[1.36]	[1.64]	[2.02]	[0.97]	[0.96]
MKTRF_t		−0.1510	−0.1010	−0.0222	−0.0240
		[−1.58]	[−1.53]	[−0.42]	[−0.49]
SMB_t			−0.5058***	−0.3193***	−0.2985***
			[−6.86]	[−5.92]	[−5.35]
HML_t			0.4663***	0.1297***	0.1667***
			[8.12]	[2.75]	[2.93]
RMW_t				0.4212***	0.4460***
				[6.72]	[6.49]
CMA_t				0.8199***	0.7839***
				[7.27]	[7.90]
UMD_t					0.0636
					[1.02]
Observations	309	309	309	309	309
F-stat		2.5010	43.4312	57.4710	52.8257