



Narrative triggers of information sensitivity

Kim Ristolainen

To cite this article: Kim Ristolainen (2024) Narrative triggers of information sensitivity, *Quantitative Finance*, 24:3-4, 499-520, DOI: [10.1080/14697688.2024.2335241](https://doi.org/10.1080/14697688.2024.2335241)

To link to this article: <https://doi.org/10.1080/14697688.2024.2335241>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 12 Apr 2024.



Submit your article to this journal [↗](#)



Article views: 399



View related articles [↗](#)



View Crossmark data [↗](#)

Narrative triggers of information sensitivity

KIM RISTOLAINEN *

Department of Economics, Turku School of Economics, University of Turku, FI-20014, Turku, Finland

(Received 5 June 2023; accepted 21 March 2024; published online 12 April 2024)

This research explores the factors contributing to information sensitivity in debt markets, focusing on the potential influences of uncertainty, economic performance, and journalist-dependent language. Building upon the foundational work of Dang *et al.* (Ignorance, debt and financial crises. Yale University Unpublished Working Paper, 2018), we analyze the mechanisms underlying the transition from information-insensitive to information-sensitive states—a shift with implications for potential financial crises. Leveraging machine learning techniques and daily data on variables such as default probability, information acquisition, and newspaper articles, we discern specific narrative triggers embedded within the news. Our analysis underscores the pivotal role of economic states and journalist language in inducing information sensitivity—a phenomenon intricately tied to different psychological thinking processes.

Keywords: Information sensitivity; Debt markets; Financial crisis; Machine learning; News data; Primordial thinking process

JEL Classifications: G01, G14, G41

1. Introduction

This research paper delves into an underexplored yet significant facet of the debt markets – the influence of journalist-dependent language, which reflects their psychological thinking processes, on the information sensitivity of these markets. Prior studies such as Dougal *et al.* (2012) have provided empirical evidence of a causal relationship between financial reporting and stock market performance. However, the intricate impact of journalist language and thinking processes on debt markets has yet to receive due attention.

Dang *et al.* (2018) advance the proposition that by their inherent design, debt markets function under the presumption of information insensitivity. This state persists as long as the cost of procuring precise information about the collateral of the debt contract surpasses the value of the information itself. When this balance is maintained, money markets operate optimally, with agents able to trade freely, unburdened by the need to obtain precise information due to a lack of concern that other agents will access detailed information on the value of the underlying collateral.

An integral element preserving debt information insensitivity is opacity. Dang *et al.* (2017) illuminate how banks strategically withhold information about their loans, thereby sustaining demand deposits in a money-like state. However, when a sufficiently negative news about the value of the

debt collateral surfaces, the debt transitions to an information-sensitive state, as the value of the collateral information now surpasses the cost of its acquisition. This shift can cause the freezing of money markets, and potentially trigger a financial crisis due to the fear of adverse selection, with quantities adjusting to zero instead of prices.

Empirical research broadly corroborates the theoretical connections between information sensitivity, information acquisition, non-price adjustments, and opacity. Nevertheless, the exact catalysts – the bad news – that instigate a shift from an information-insensitive state to an information-sensitive one, and vice versa, remain underexplored. This is where our research contributes.

In this paper, we do not merely identify the narrative triggers that induce this transition to information sensitivity; we shed light on how the language employed by journalists, indicative of their thinking processes, can greatly influence whether a topic serves as a trigger. Using a machine learning algorithm and daily credit default swap (CDS) spreads and Google search data as proxies for default probability and public information acquisition about a firm, we discern distinct states categorized by these two variables. These states, which we characterize as information (in)sensitive, are labeled based on their respective firm-day observation, and we further delineate the days when a shift to either state has occurred.

To study the general factors prompting these state shifts, we combine the daily data on the information sensitivity

*Corresponding author. Email: kim.ristolainen@utu.fi

states of 576 financial and non-financial companies with news article data from the Wall Street Journal. By utilizing natural language processing and machine learning techniques, we identify 80 latent topics and their daily frequencies from 1890 to 2022. We then proceed to extract the unexpected attention to each news topic on a given day, defining unexpected attention as the part of news topic prevalence that was unpredictable based on past news attention data.

We use local projection regressions (Jorda 2005) to uncover several topics that, when they receive increased attention, trigger a rise in the probability of companies transitioning to an information-sensitive state post-news publication. However, this narrative trigger effect varies significantly with the state of the aggregate economy or the specific firm and the language used by the journalists. The journalist language differences reflect their individual thinking processes as assessed through the application of Martindale's (1975) regressive imagery dictionary to the primary-conceptual thinking processes continuum introduced by Freud (1938). Although some journalists do not consistently lean towards either thinking processes, there exist distinct clusters of journalists who regularly utilize language associated with one of the thinking processes throughout their careers. This difference in thinking process as expressed in language has a profound influence on the effectiveness of the narrative triggers.

Our findings contribute to the empirical research on information sensitivity. We introduce a novel approach to measuring an individual firm's daily information sensitivity state and illuminate the general triggers of information sensitivity that can be quantified at a daily frequency. Moreover, our research demonstrates that both aggregate and idiosyncratic uncertainty, as well as economic performance and journalist-specific language, play roles in determining whether a topic serves as a trigger. This underscores that the dynamics of information sensitivity are far from purely mechanical, but are strongly influenced by the human factors in financial journalism.

The remainder of the paper is organized as follows: section 2 focuses on identifying the daily information sensitivity state for individual firms. Section 3 outlines the creation of a measure for unexpected attention to news topics from text data. Section 4 delves into the journalists' thinking processes. Section 5 presents the empirical results for information sensitivity triggers. Finally, section 6 concludes our findings.

2. Identifying information sensitivity

Dang *et al.* (2018, 2020) define the concept of information sensitivity in the following way: An agent can buy a security with price p and payoff $s(x)$, where the random variable x has a probability density function $f(x)$. The agent can produce information about the exact value x at a cost γ . The authors define the value π_L of producing private information when the agent perceives the security as undervalued (price p is lower than the expected value of the payoff $E[s(x)]$), as the potential loss that would occur if the payoff $s(x)$ were to be smaller than the price p of the security. More formally, the value of

information or the information sensitivity in the loss region is

$$\pi_L = \int \max[p - s(x), 0]f(x)dx \quad (1)$$

In the case where the agent perceives the security as overvalued ($p > E[s(x)]$), the value of information, π_H , is the expected loss if the agent does not buy the security and, in fact, p happens to be smaller than $s(x)$. More formally,

$$\pi_H = \int \max[s(x) - p, 0]f(x)dx. \quad (2)$$

The authors show that the information sensitivity (value of information) of a security to the buyer or the seller is $\pi = \min[\pi_L, \pi_H]$ for any p and $f(x)$.

To make the decision about producing private information, the agent will assess whether the value of information, π , is higher than its cost, γ . When no agent deems acquiring information as profitable and all agents are (rationally) aware of this, the security is seen as information-insensitive. Dang *et al.* (2018) show that debt is the most information-insensitive security and that when it is backed by debt, the information insensitivity is maximized. They also argue that debt is inherently vulnerable to crisis in the sense that when a sufficiently large negative shock related to the value of the collateral backing the debt occurs, there is a positive probability that there will be no trade at all, as some investors can and will produce private information while others are deterred by the fear of adverse selection.

Information sensitivity has been extensively studied from an empirical perspective, focusing on the predictions made by the theories of Dang *et al.* (2018). These empirical studies have confirmed several key aspects of information sensitivity, including the increase in information production about debt collateral during an information-sensitive state (Brancati and Macchiavelli. 2019, Gallagher *et al.* 2020), the adjustment of debt quantity rather than price in response to bad news (Gorton 1988, Perignon *et al.* 2018), and the impact of opacity and transparency on information insensitivity (Baghai *et al.* 2022, Cipriani and La Spada. 2021). While these studies have empirically validated the existence and characteristics of information sensitivity, they have not examined the specific triggers that lead to state switches. Identifying these triggers is crucial for a deeper understanding of the dynamics of harmful events associated with information sensitivity.

To identify and examine potential triggers of information sensitivity empirically, we need to measure the information sensitivity state of a firm and potential trigger candidates over time. Previous empirical studies have demonstrated the presence of information sensitivity through significant effects in regression frameworks that test the relationship between key variables predicted by the theories of Dang *et al.* (2018). In our analysis, we go beyond these studies by not only utilizing the predicted relationship between information production and bad news but also labeling each company-day with a specific information sensitivity state based on the firm's default probability measured by CDS spreads and the public's information acquisition measured by Google searches.

Based on the characteristics of the information sensitivity property, we hypothesize four possible states in the default

probability (DPR)—public information acquisition (PIA) space of a firm’s debt. These states include:

1. **Information-Insensitive State:** This state corresponds to days when information is not acquired, and the default probability of a company remains low. The company can be considered in an information-insensitive state during these periods.
2. **Trending State:** In this state, information about a company is acquired, but the default probability remains low, suggesting that the company is trending due to factors unrelated to default risk. The company’s debt is still considered information insensitive during this state.
3. **Information-Sensitive State:** When a company’s debt becomes information sensitive, significant information acquisition occurs, accompanied by a rapid increase in the default probability.
4. **Default State:** This state occurs when a company is no longer trending but has a very high and relatively stable CDS spread, indicating a high default probability.

Our objective is to categorize each company-day observation into one of these states for further analysis.

2.1. Gaussian mixture model

To identify the different information sensitivity states, we employ a Gaussian mixture model (GMM), a popular choice for mixture models that has been used in various fields such as modeling stock returns (Kon 1984, Malevergne *et al.* 2005, Behr 2007). The GMM assumes that each state m follows a multivariate normal distribution with its own mean μ_m and covariance matrix Σ_m . Formally, the GMM can be expressed as:

$$f(\mathbf{x}) = \sum_{m=1}^M \theta_m g(\mathbf{x}; \mu_m, \Sigma_m), \quad (3)$$

where \mathbf{x} represents the observed variables (CDS spreads and Google search data), M denotes the number of states, θ_m represents the mixing proportions, and $g(\mathbf{x}; \mu_m, \Sigma_m)$ denotes the multivariate normal distribution. The unknown parameters, including the mixing proportions, means, and covariance matrices, are estimated using the expectation-maximization (EM) algorithm. The EM algorithm optimizes these parameters to maximize the log-likelihood given by equation (3). In the EM algorithm, initial guesses for the unknown variables are set, and then an iterative process of expectation and maximization steps is performed until convergence. The responsibilities or conditional expectations of observations belonging to specific states are calculated in the expectation step, and updated values for the unknown parameters are obtained in the maximization step. The process is repeated until convergence is achieved (Hastie *et al.* 2009).

The number of components or states, M , needs to be pre-determined before estimating the unknown parameters of the model. Since there is no definitive method for determining the optimal number of states, a common approach is to select the model that maximizes the increase in the Bayesian information criterion. In our analysis, we choose to have four

components, as this number is likely to capture the simplest model that can approximate the hypothesized states.

2.2. Characteristics of information sensitivity states

We collect the available 5-year CDS spread data from Refinitiv Datastream, including both non-financial and financial companies, for the period 2006–2022. The CDS spread serves as a proxy for the default probability of a company. To measure the public’s information acquisition (PIA), we gather daily Google trend data, which approximates the level of information acquisition related to specific companies.† By merging the CDS and Google trend data, we construct a panel of matched daily observations for both variables, resulting in a dataset of 576 companies and over 1.9 million daily observations.

The estimated values of the unknown parameters based on our extensive dataset are presented in Panel A of table 1. These results support our hypothesis that there are four distinct states: information-insensitive state, trending state, default state, and information-sensitive state. The table provides the means, standard deviations, sample sizes, and share percentages for each state.

To examine the persistence of each state, we report the conditional probabilities of a firm being in a specific state at period t given the state in the previous period $t-1$ in Panel B of table 1. Despite the model having no temporal information, the states exhibit strong persistence, as the majority of observations maintain the same state from the previous period. Additionally, the results align with the assumed evolution of information sensitivity, with the default state most commonly following an information-sensitive state, and the information-sensitive state frequently preceding the information-insensitive state. Notably, transitions directly from the default state to an information-insensitive state or vice versa are extremely rare. The most prevalent states in the dataset are the information-insensitive state and trending state, accounting for approximately 66.1% of the firm-days. The default state is the least common, comprising only 3.9% of the observations, while the information-sensitive state represents nearly 30% of the total observations.

The evolution of CDS spreads for six non-financial and six financial corporations from 2008 onwards, classified according to their information sensitivity states, is depicted in figures 1 and 2. The model effectively captures shifts from calm to turbulent periods without assigning incorrect labels in the midst of a particular state. Furthermore, we provide specific examples of state switches for several companies.

For instance, Macy’s experienced a challenging year in 2015, with a decline in sales during the second half. Our measure indicates a clear switch to information sensitivity on November 9, 2015, two days before the release of the company’s disappointing third-quarter earnings. While the sales

† The daily trend data from 12/2007 to 2/2022 are collected by first downloading the monthly data for the entire search period for the search term and then gathering the daily data per month. The daily data are then made comparable between months by multiplying the daily data by the monthly search volume and dividing by 100 to account for Google’s smoothing of large trend requests.

Table 1. Information sensitivity states of companies.

Panel A	Variable	Mean	SD	N	Share %
Trending for other	CDS spread	72.9	34.9	941 001	48.9
	Google searches	36.1	22.9	941 001	48.9
Insensitive state	CDS spread	101.7	64.0	331 201	17.2
	Google searches	0	0.1	331 201	17.2
Default state	CDS spread	3259.5	3500.7	75 916	3.9
	Google searches	14.5	18.5	75 916	3.9
Sensitive state	CDS spread	307.5	178.4	575 884	29.9
	Google searches	28.9	22.9	575 884	29.9

Panel B	Trending for other _{t-1}	Insensitive state _{t-1}	Default state _{t-1}	Sensitive state _{t-1}	Total share %
Trending for other _t	90.986	8.525	0	0.463	100
Insensitive state _t	24.22	69.076	0	6.661	100
Default state _t	0	0	97.946	2.022	100
Sensitive state _t	0.747	3.836	0.267	95.123	100

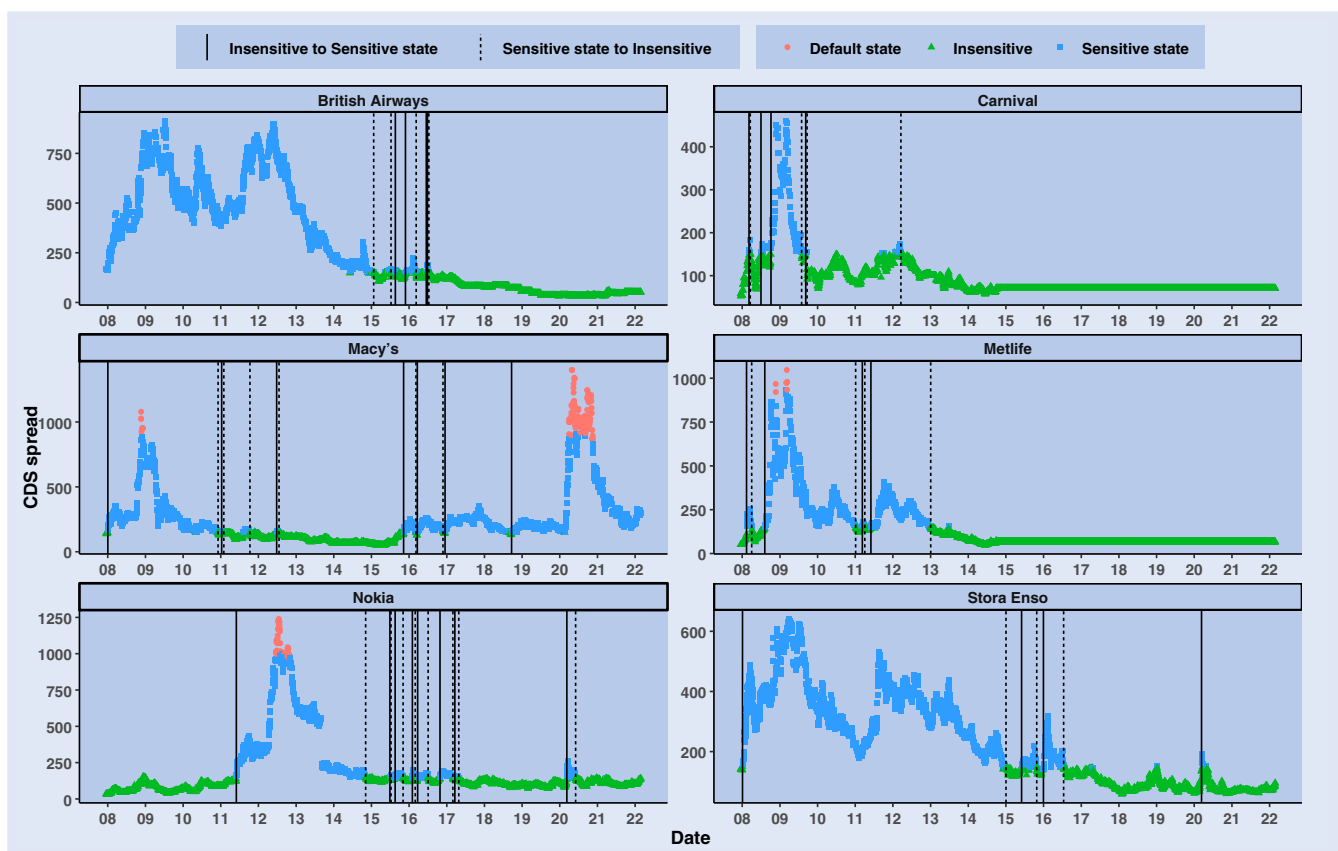


Figure 1. Information sensitivity states and CDS spreads for non-financial firms.

drop of 5.2% reported by the company (FT.com 2015) contributed to the switch, it is likely that the news from financial analysts revising their price targets for retail companies due to excess inventory and unusually warm weather (Kapner 2015) played a significant role.

Another notable switch to information sensitivity occurred on June 1, 2011, for Nokia. The previous day, the company issued a profit warning primarily due to the increasing success of phones using Android operating systems in the European market (Lawton and Efrati. 2011). Carnival, a cruise vacation company, declared on October 31, 2008, that it would suspend dividend payments to bolster its cash reserves and reduce reliance on capital markets (Curran 2008). Although

stocks reacted negatively to this announcement, our measure indicates that the firm became information sensitive almost four weeks earlier on October 6.

Figure 3 displays the evolution of the number of companies that are in an information-sensitive state in specific periods. The significant fluctuations in the number of information-sensitive companies appear to capture events related to financial turmoil and stability. For instance, this number declined following the Federal Reserve's emergency meeting on March 14th, 2008, regarding Bear Stearns. Similarly, a decrease occurred after the Centre-right party narrowly won the Greek elections on June 17th, 2012, and when ECB President Mario Draghi delivered his famous 'whatever it takes

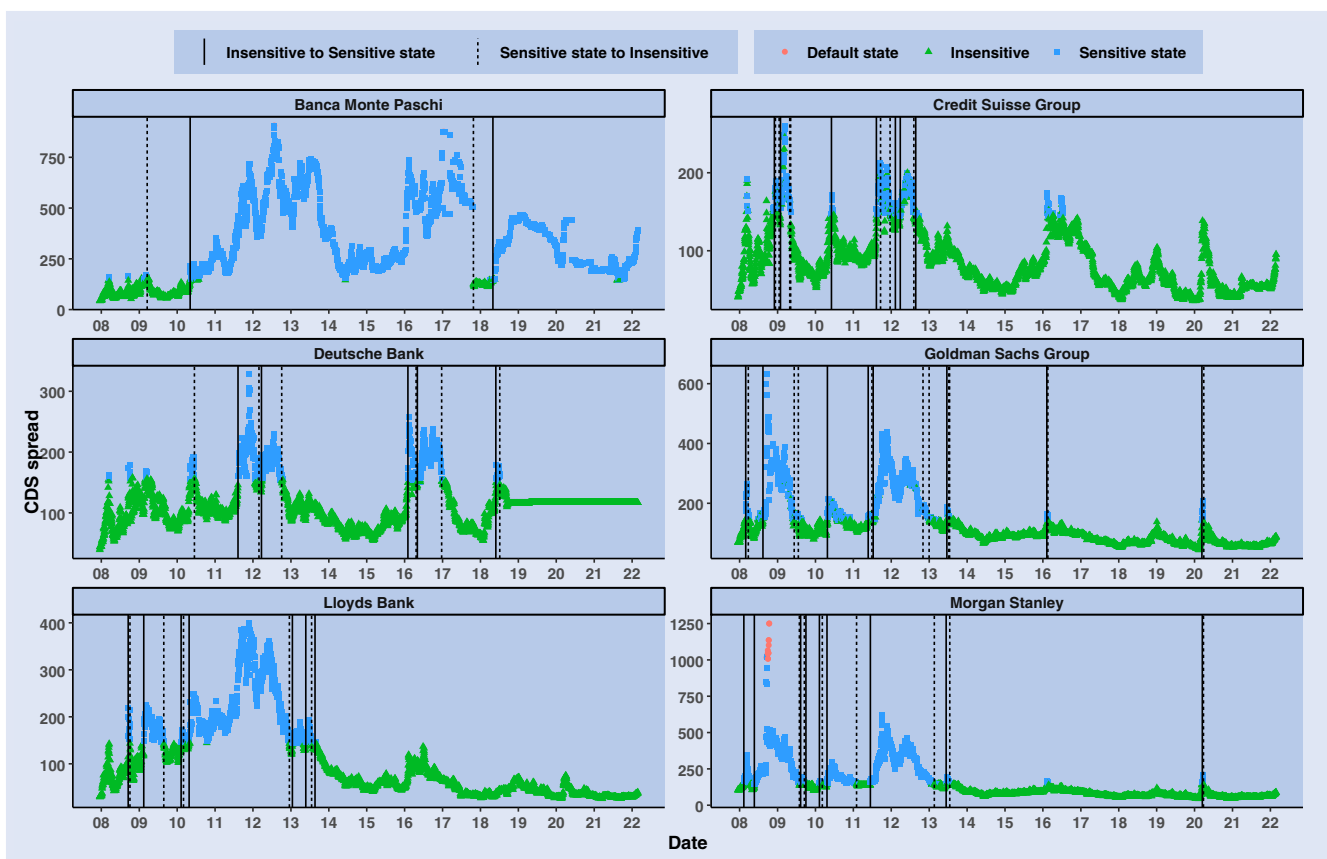


Figure 2. Information sensitivity states and CDS spreads for financial firms.

to preserve the euro’ speech on July 26th, 2012. In contrast, the number of information-sensitive companies in the sample increased sharply following Lehman Brothers’ bankruptcy filing on September 15th, 2008, and the onset of the Covid-19 pandemic in February 2020.

3. Measuring news surprises

In the preceding section, we demonstrated that our information sensitivity measure effectively captures the timing of transitions between distinct and discernible states associated with a company’s default risk and the public’s interest in gathering information about the company. While our measure aligns with known instances of information sensitivity state switches, our goal is to identify news content that serves as a general trigger for information sensitivity and can be quantitatively assessed. To accomplish this, we need to first identify common content patterns in historical news articles and measure the prevalence of specific content within a given time period or individual news titles.

3.1. Attention to economic news topics in 1890–2022

To measure the attention that a news topic had on a specific day between 1890 and 2022, we estimate an extension of the most commonly used topic model, Blei *et al.*’s (2003) latent Dirichlet allocation (LDA) model. Topic models are unsupervised learning models that try to uncover latent topics from a

collection of text documents. These models assume that each text in the corpus is generated by a specific generative process. In the LDA, each text document d can consist of multiple topics k , and each topic k has a word distribution β stating how likely it is to observe a specific word from the fixed vocabulary V that holds all the unique words found in our text corpus. In addition, each document d has a topic distribution θ_d that represents the proportions of each topic that the document consists of.

The generative process works in the following way. First, a topic assignment $z_{n,d}$ is generated for each word position n for each document d from the topic distribution θ_d . Then, a word assignment $w_{n,d}$ is generated from the word distribution β_z given the topic assignment $z_{n,d}$. Both β and θ are assumed to be distributed according to a Dirichlet distribution with parameters α and η . These parameters influence how focused the Dirichlet distribution is either on the middle (documents with multiple topics) or on the corners of the distributions (documents with few topics). More formally, with a corpus of M documents with N words and K topics, the probability of observing a corpus can be written as follows:

$$P(\theta, \beta, Z, W) = \prod_{k=1}^K P(\beta_k | \eta) \prod_{d=1}^M P(\theta_d | \alpha) \prod_{i=1}^N P(z_{d,n} | \theta_d) P(w_{d,n} | \beta, z_{d,n}). \tag{4}$$

Given the word assignments $w_{d,j}$ and the number of topics K , the unknown parameters are estimated with Gibbs sampling.

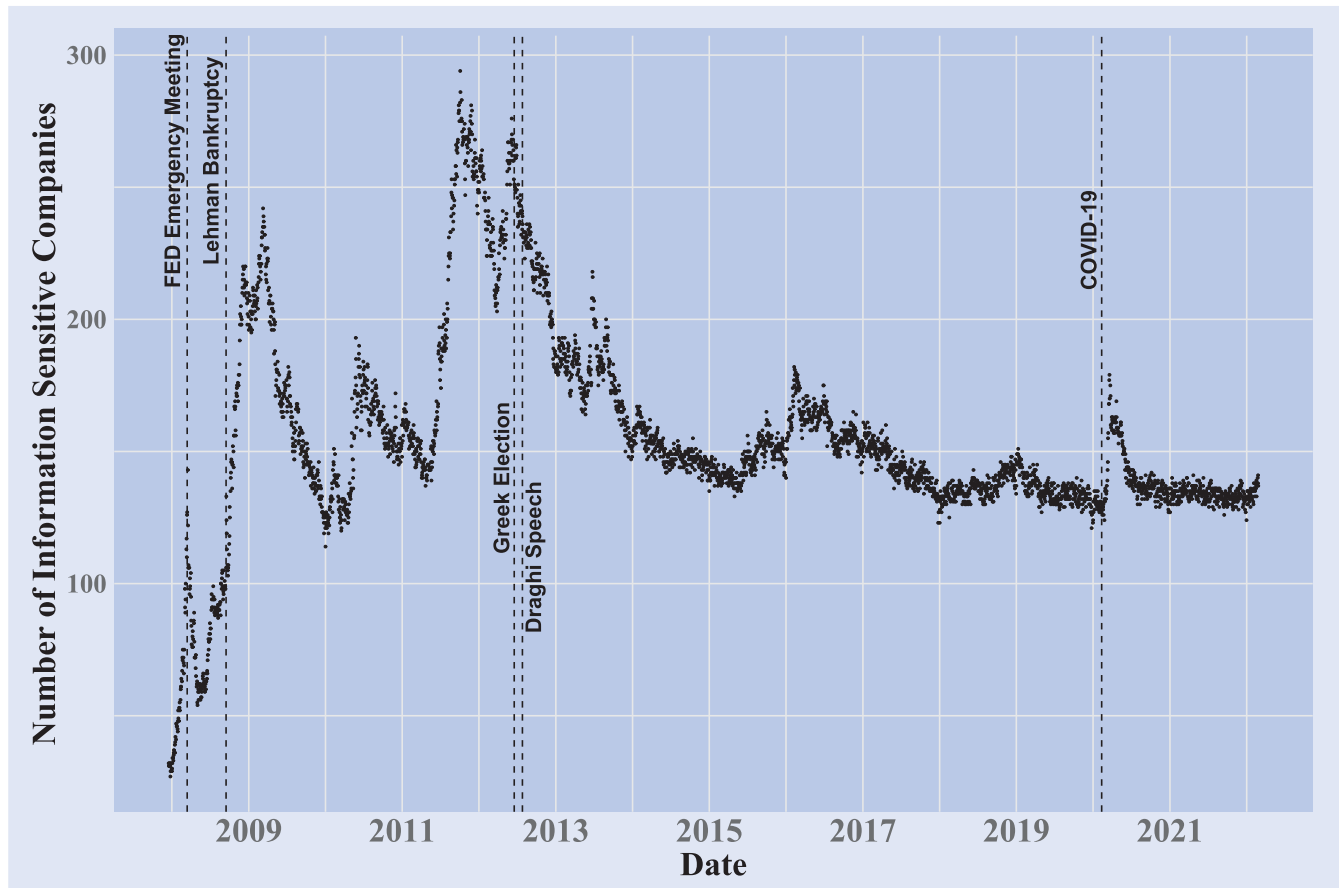


Figure 3. The evolution of the number of information sensitive companies in time.

Note: The following events are displayed with vertical lines. *FED emergency meeting in March 14th, 2008:* The Federal Reserve board had an emergency weekend meeting regarding Bear Stearns. *Lehman Bankruptcy in September 15th, 2008:* Lehman Brothers filed for bankruptcy. *Greek election in June 17th, 2012:* The centre-right wins legislative elections in Greece. *Draghi Speech in July 26th, 2012:* ECB president Mario Draghi gives the famous ‘the ECB is ready to do whatever it takes to preserve the euro’ - speech. *COVID-19 in February 11th, 2020:* WHO names the new virus as COVID-19.

An important limitation of the LDA is that it assumes the topics to be uncorrelated. This is a relatively unrealistic assumption, as observing a specific topic in a document might give us information that it is likely to discuss topics that are related to the observed topic rather than completely unrelated topics. For example, if the corpus included lifestyle magazines, then if we observe a car topic without knowing that it is in a men’s magazine, we would think that it is more likely to also have content about sports rather than women’s fashion in the magazine. To account for this issue, Blei and Lafferty. (2005) introduced the correlated topic model (CTM), which allows topics to be correlated. The CTM generative process differs from that of the LDA. The topic distributions θ_d are not from a Dirichlet distribution, but they are distributed according to a logistic normal distribution with a mean μ and a covariance matrix Σ with K dimensions. The covariance matrix enables the model to capture correlations between topics.

We estimate the CTM with a corpus that includes the titles of all news articles published in the *Wall Street Journal* in the period 1890–2022. The text data were gathered from Proquest Historical Newspapers using their text and data mining (TDM) tool. The news titles were cleaned[†] before they were

transformed into a numerical format as data feature matrices (DFMs) that are used as inputs in a topic model. Each element of a DFM represents a word count, where the rows correspond to individual documents, and the columns represent unique words found in the corpus. We select the optimal number of topics with Mimno and Lee.’s (2014) algorithm. This algorithm utilizes the assumption that each topic has a specific anchor word that appears only in that specific topic. The authors show that using their algorithm to find the anchor words and then using these words in the estimation process of the topic model results in better topics quantified with many different measures.

Table A1 presents the 80 topics of the estimated topic model with the labels and the most common words of each topic. The majority of the topics are highly identifiable from the most common words and are also quite separable from other topics. This can be seen in the fan dendrogram of figure 4, which visualizes the topics with a hierarchical clustering algorithm that uses information from each topic’s word distributions e.g. topics whose vocabulary is more similar are more likely to be grouped together. The model seems to capture a vast spectrum of different topics found in economic news in the past 130 years, ranging from insurance, debt markets, inflation, financial regulations and banking to natural disasters, crime, court rulings, political campaigns, military,

[†] This process is described in detail in the Appendix.



Figure 4. Hierarchical clustering of topics. The dendrogram plots the result of a hierarchical clustering model estimated with the topic word-distributions.

wars and diseases. The model also identifies topics that are likely irrelevant for the economy, such as food, family, music, art, design and sports. Finally, the heatmap of topic prevalence in figure 5 visualizes economic news reporting over time.

3.2. Unexpected news content

The output of the topic model that we want to utilize is the topic-word distributions β_k for each topic k and the document-topic distributions θ_d for each news article title d . The former can be used to label the topics and the latter to see which topics a specific new title consists of. We further aggregate the topic distribution information to a daily topic attention series by averaging the share of each topic for each day for the entire

time period. As we are interested in the possible triggers of information sensitivity switches, we prefer to have a measure of news that enables us to state more about causal relationships, not just correlations. Therefore, we form measures of unexpected attention to different news topics. Unexpected attention means that this attention could not have been foreseen with prior information. This type of measure captures, for example, the start of the sudden increase in disease and medication news due to the COVID-19 pandemic, but then quite quickly normalizes after the beginning of the reporting, as then the attention to that topic is no longer a surprise.

In their work, Glasserman and Mamaysky. (2019) construct a metric for identifying unusual news. They calculate the uncommonness of a specific n-gram in a given period t

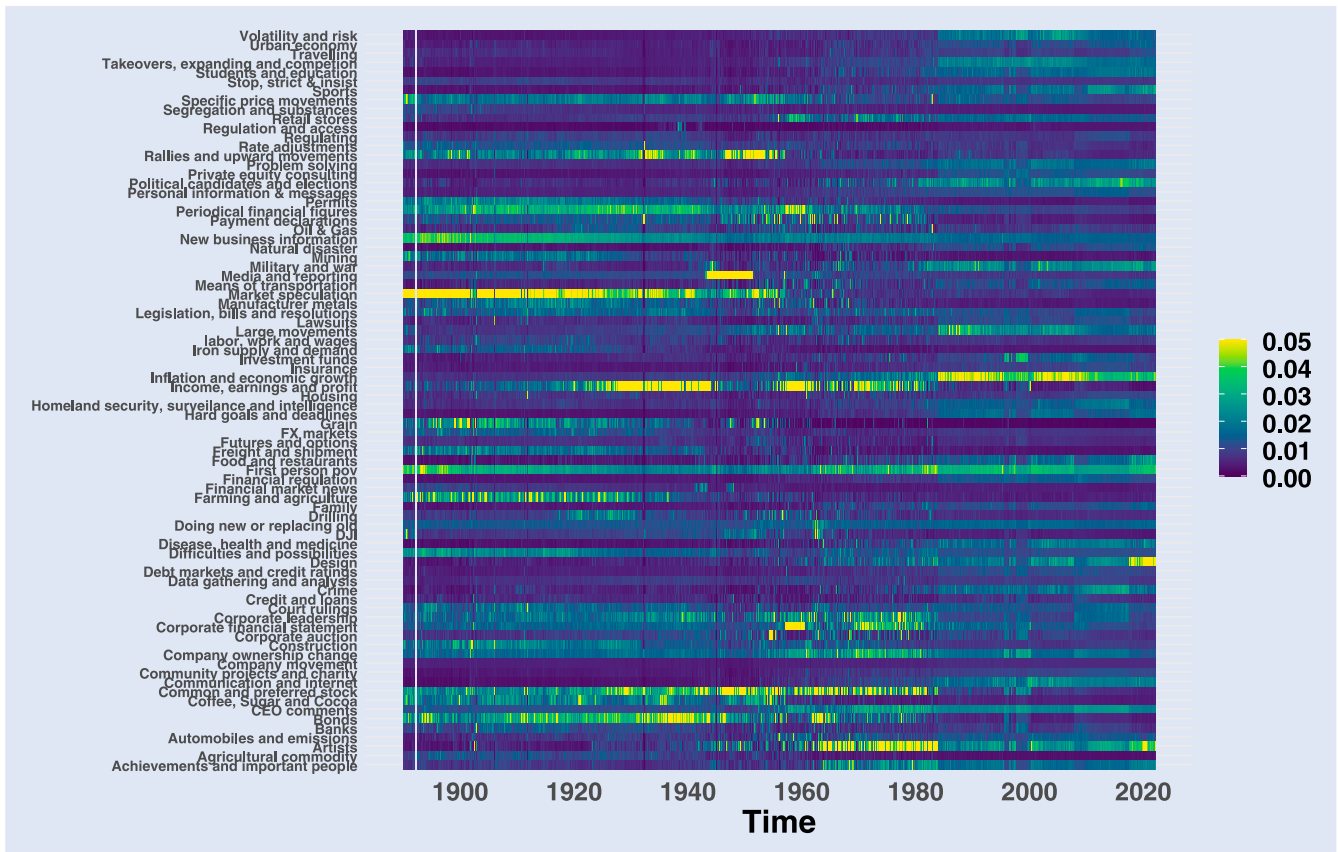


Figure 5. Prevalence of news topics in Time. The figure plots the topic distributions of each topic k aggregated to a monthly level across the period 1890–2022.

by comparing its actual frequency in that period against its expected frequency, based on a historical corpus of news prior to t . This individual n-gram measure is then aggregated to gauge the overall unusualness of a text at t . Specifically, a text is considered unusual if it contains n-grams that are prevalent in current news but scarce in historical news. Our approach shares similarities in capturing unexpectedness based on historical news context. However, our primary distinction lies in emphasizing the salience of topics that comprehensively represent the news content, rather than focusing solely on a generalized metric for the unusualness of news for a specific period t .

Our approach to measure the unexpected share of a given topic in the news is very similar to the procedure that Bianchi *et al.* (2022) used to extract biases in people’s beliefs. The authors estimate a machine learning model with available objective information up to that point in time to get a benchmark prediction for the same statistic that survey respondents are also predicting. With this procedure, one can analyze what one predicted and what one should have predicted given available public information. We utilize this procedure for a different task by using it to get an objective prediction for a news topics prevalence in news reporting given recent and historical trends in news reporting. Next, we discuss in detail how we extract unexpected news from the news topic data.

First, we estimate the expected topic proportions for each topic for each day given the information on past news. This is done so that a flexible elastic net model is estimated with cross-validation to predict tomorrow’s topic distribution,

given the information on past topic distributions of the last 5 years. Then, an out-of-sample prediction is made for the next day’s topic distribution. The out-of-sample prediction error is used to measure the unexpected share of attention each topic has on a given day. The procedure can be presented in the following way:

- i. An elastic net model (Zou and Hastie. 2005) is estimated to predict the average share $Y_{k,t}$ of topic k in the news on day t with information X_{t-1} about all topic distributions† up to day $t - 1$. The elastic net model can be formally presented as

$$\min_{\beta_0, \beta} \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \beta X_i)^2 + \lambda \sum_{j=1}^P \left(\frac{(1 - \alpha)}{2} \beta_j^2 + \alpha |\beta_j| \right),$$

where λ is a regularization parameter that determines how much shrinkage and sparsity are introduced to the model via Ridge regression and least absolute shrinkage and selection operator (LASSO) penalties. The optimal value for λ is estimated with 5-fold

† The predictors X_{t-1} include the mean topic proportions of the previous 3 days ($t - 3$ to $t - 1$), and the mean and the standard deviation of the topic proportions of the previous week ($t - 8$ to $t - 1$), month ($t - 30$ to $t - 1$) and 6 months ($t - 180$ to $t - 1$) for all K topics, implying a total of 720 predictors with 80 different topics.

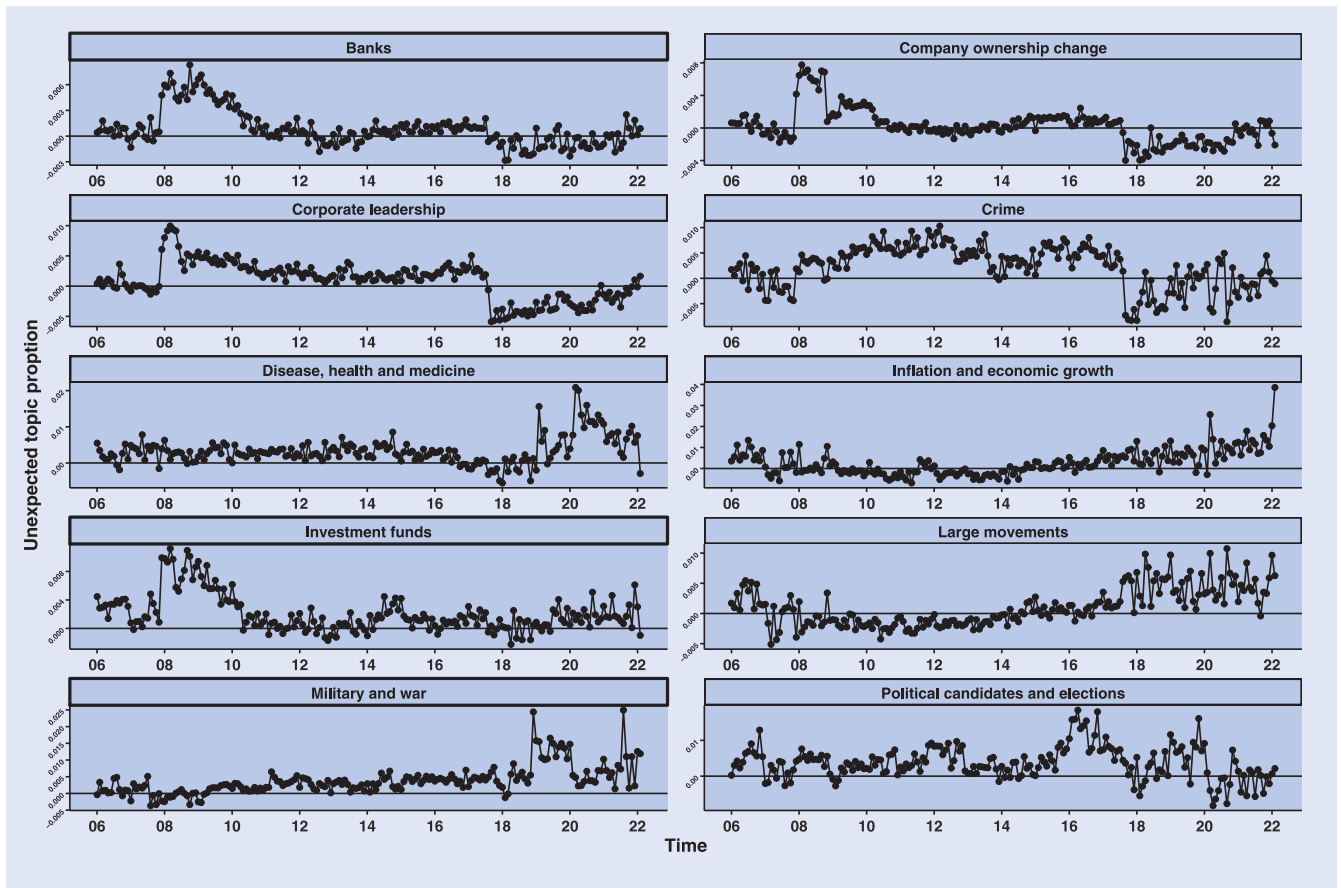


Figure 6. Evolution of unexpected news in selected topics. The figure plots the unpredictable part of topic's daily prevalence for each topic aggregated to a monthly level across the period 2006–2022.

cross-validation, where each 20% proportion of data is reserved once as a validation set, and the model is estimated with the remaining 80% of the data. The prediction error for the validation set is collected, and the λ that minimizes the average mean squared error (MSE) for these five validation errors is chosen as the optimal one. The model is estimated with the data from the previous 5 years.

- ii. Step i is repeated for each day t and topic k in a rolling window fashion to get an out-of-sample prediction for the topic proportion in period t , with an elastic net that was estimated with data available only before period t .
- iii. Finally, to extract the unpredictable part of the attention to a topic, we collect the out-of-sample prediction error for each topic k for each day t .

To clarify, our purpose is not to measure whether a specific news title or event was completely unexpected, but whether the daily attention to a specific general topic was unexpected. Because the unpredictable part of the attention to a topic is independent from past information, we have a measure of a shock in news attention to a specific topic.

The daily unexpected news attention series for a group of selected topics aggregated to a monthly frequency is plotted in figure 6. The measure seems to work well as it captures some highly significant and unexpected shifts in news reporting. The start of the global financial crises of 2008 can be

clearly seen in the figures, as the banks, corporate leadership, investment funds and company ownership topics receive more unexpected attention in the news during those periods. In addition, the political candidates and elections topic receive unexpectedly large attention during the 2016 and 2020 U.S. elections relative to previous elections. The disease, health and medicine topic seems to peak in early 2020 when the COVID-19 pandemic began. In addition, the inflation and growth topic surprisingly receives much attention in 2022, when inflation started to rise astonishingly fast.

Columns 3 and 4 of table A1 report the out-of-sample mean absolute errors (MAEs) for the prediction of each topic by the elastic net model and the share of positive surprises in attention to each topic for the entire time span. It seems that it is more common that an increase rather than a decrease in attention to a specific topic is unexpected. The results imply that some topics are, in general, clearly more unpredictable than others. For example, the attention to commodity, agricultural, exchange rate, manufacturing material and work, labor and wages is much more predictable than the attention to large movements, research and education, disease, health and medicine, military and war, and inflation and growth topics. This makes sense, as specific topics often relate to periodical and seasonal reporting and events, and others are more unpredictable by nature. With these observations, we infer that our measure captures the unexpected attention to news topics sufficiently.

4. Journalists' thinking processes

News is supposed to be an objective source of information about events of different levels of importance. However, the writing style, creativity and language used can vary across journalists, and even among articles written by the same journalist. In addition to the news content, these aspects of the text can affect the signals that the economic agents receive from news articles. This variation in writing style can be a result of external (mood and other personal events) and news content-related (journalist subjective opinion/view about the news and its possible effects on the world) factors specific to the journalist. The meaning of news content varies across reporters. Psychological literature explains why a journalist's personal relationship with the news content can materialize in the way the news article is written.

Freud (1938) argued that a person's personality consists of the id, the ego and the superego. The id is seen to be the most primitive part of the personality, and it is the first part of the personality that evolves when a human is born. According to Freud, the so-called primary thinking process is a way for the id to handle the primitive urges that the pleasure principle creates. When a person grows older, the ego and the superego play a larger role in a person's personality, and the secondary or conceptual thinking process emerges to tackle the urges to satisfy primary needs that are not suitable in the real world. These two thinking processes were introduced in the psychological literature by Freud (1938) and further discussed in Goldstein (1939) and Werner (1948).

The primordial or primary thinking process has been seen to relate to thinking that is irrational, free-associative, sensational, impulsive, concrete and unconcerned with a purpose. Primordial thinking is thought to be free of time, space, real world and social institutions; thus, it is more common during dreams, fantasy and the use of drugs. On the other hand, conceptual or secondary thinking is rational, reality-oriented, problem solving, logical, conceptual and narrowly focused (Svensson *et al.* 2006, Granger 2011, Kopcsó and Láng, 2019). Primary thinking has been associated with creativity (Martindale 1998). Katz (1997) argued that the primary process is used during the inspirational, incubation and illumination phases of the creative process, whereas the conceptual thinking process is used later during a verification phase. Journalists' primary feelings related to a news event might trigger the primary process during the writing process and emerge as a specific type of language used in the text. For example, a journalist might have strong feelings or opinions about specific politics, laws, or natural disasters that span from her id that developed early in her childhood. There might be a primary need to react to the news content, and the journalist's primary process facilitates this urge during the writing process.

To measure a journalist's mental thinking process, we utilize the regressive imagery dictionary developed by Martindale (1975). The dictionary is a collection of words that are seen to relate to either primordial or conceptual thinking. Many papers have validated this dictionary by showing that primary process words are more common in written text during coprolalic verbal ticks symptoms of people with Gilles de la Tourette's syndrome (Martindale 1977), during

the use of marijuana (West *et al.* 1983), in stories that are more creative (Martindale and Dailey, 1996) and among people who are writing in the dark and suffer from the fear of dark relative to texts written in well-lit areas (Kopcsó and Láng, 2019). The words of the thinking processes can be further divided into different subcategories. Examples of the subcategories of primary thinking words are vision, concreteness, unknown, brink passage, general sensation, hard, soft, consciousness alteration, diffusion, narcissism, concreteness, passivity, voyage, random movement, chaos, timelessness, diffusion, touch, taste, odor, sound, cold and conscious. Secondary process words are about abstraction, social behavior, instrumental behavior, restraint, order, temporal references and moral imperatives (Martindale 1977).

As primary process thinking is related to specific aspects, such as creativity, impulsiveness, irrationality etc., the share of primary and secondary thinking processes words among the texts in news articles discussing the economy and companies whose debt the agents hold (or whose debt is the collateral for the debt they own) can give signals that distort, emphasize, diminish, magnify, raise doubt, confuse or elucidate the message about the fundamental content of the news. In addition, the primary thinking process can emerge from agents who are the subjects of the news. For example, there were a lot of different ways that Mario Draghi could have given the message in his famous speech on July 26, 2012. If he had left out the phrases *the ECB will do whatever it takes* and *you better believe it is enough* from the speech, then the message may not have been as persuasive, and the European debt markets might have remained in turmoil.

We measure the thinking process continuum TP_d behind document d in the manner common to the literature (Martindale *et al.* 1986, Martindale 2007, Kopcsó and Láng, 2019) as the difference between the shares of primordial thinking process words and conceptual thinking process words.† More formally,

$$TP_d = \text{Primary words share } \%_d - \text{Conceptual words share } \%_d. \quad (5)$$

We aggregate this measure to daily TP_t and author-level TP_a measure in the following way:

$$TP_t = \frac{1}{N^t} \sum_{d \in t} TP_d, \quad (6)$$

$$TP_a = \frac{1}{N^a} \sum_{d \in a} TP_d. \quad (7)$$

This measure captures in which direction on the primordial–conceptual thinking process continuum the news article texts lean. Different statistics characterizing TP_a across authors‡

† As the two thinking processes and the language related to them are seen as opposites of each other, the thinking process language continuum is often measured as the difference or the ratio of the shares of primordial and conceptual words.

‡ We include authors who have written at least one article since 1.1.2006 as this is the first date for which we also have data on CDS spreads, Google trends and hence information sensitivity that we use in the main analysis in Section 5.

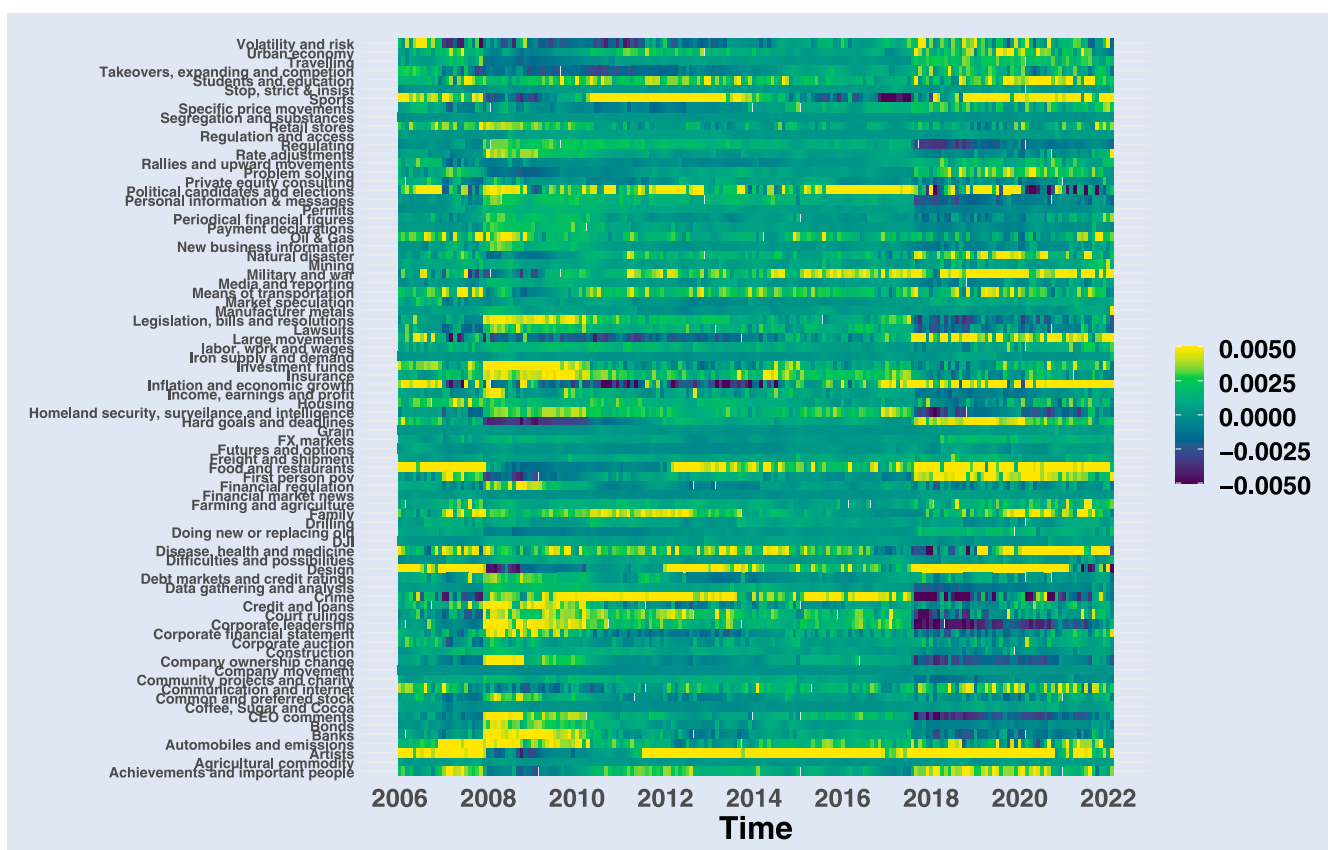


Figure 7. Evolution of unexpected news in time. The figure plots the unpredictable part of topic's daily prevalence for each topic k aggregated to a monthly level across the period 2006–2022.

are plotted in figure 8. Figure 8(a–c) reveal that the clear majority of journalists use the conceptual thinking process more on average, but interestingly there is also a relatively large group of authors who lean more to the usage of primary thinking process language on average. However, the large dispersion in author-specific standard deviations implies that the thinking process is in no way constant and varies substantially for each author and more, among others. Interestingly, for a large share of authors, the thinking process is very persistent (a positive auto-correlation) and for the large majority, it is not that persistent. There are also authors whose thinking process across time has a negative autocorrelation, implying that they switch persistently to the other process after each news article. This descriptive evidence point to the fact that these two thinking processes are present in news articles.

Figure 9(a) plots the 25th and 75th quantiles and the mean of TP_t for each year from 1890 onwards.[†] It appears that significant shifts in the shares of conceptual and primary thinking process language occurred in the news throughout this period. These shifts are characterized by decade-long gradual increases or decreases in the ratio of primary to conceptual thinking process language. The most significant increases were observed in the 1890s, 1940s, 1960s, and 1980s. Rapid and sustained increases were evident in the early 2000s and late 2010s. Conversely, the most pronounced decreases occurred in the 1950s and 1990s. There were also

[†] Figure 8(a,b) include the statistics for texts written by authors represented in Figure 7(a–c) and also for the texts where author information was not available.

sharp, relatively sustained declines in the late 1960s and around the global financial crisis of 2008. Notably, this share remained relatively consistent from the early 1900s up to the onset of World War II. It is also plausible that these different thinking processes are more common in some topics than in others. Figure 9(b) displays the monthly correlation of a topic's prevalence and TP_d across topics and time. What is striking is that although there is variation in the correlation across topics, it seems to be high during specific longer time periods. For example, the language in news articles was clearly leaning toward the conceptual thinking process in the 10 years following the Second World War. In addition, the primordial thinking process was relatively more present from 1955 to 1985.

5. Triggers of information sensitivity

5.1. Systemic triggers and company specific attention

To investigate the influence of unexpected attention to different news topics on companies' information sensitivity, we employ the local projection method of Jorda (2005). We estimate the following specification:

$$\Delta_h Y_{i,t} = \alpha_i^h + \beta_k^h \sum_{k=1}^{80} A_{k,t} + \omega_k^h \sum_{k=1}^{80} F_{i,t,k} + \gamma_k^h \sum_{k=1}^{80} A_{i,k} F_{i,t,k} + \eta^h Z_t + \zeta^h X_{i,t} + \epsilon_t \quad \text{for } h = 1, \dots, 30. \quad (8)$$

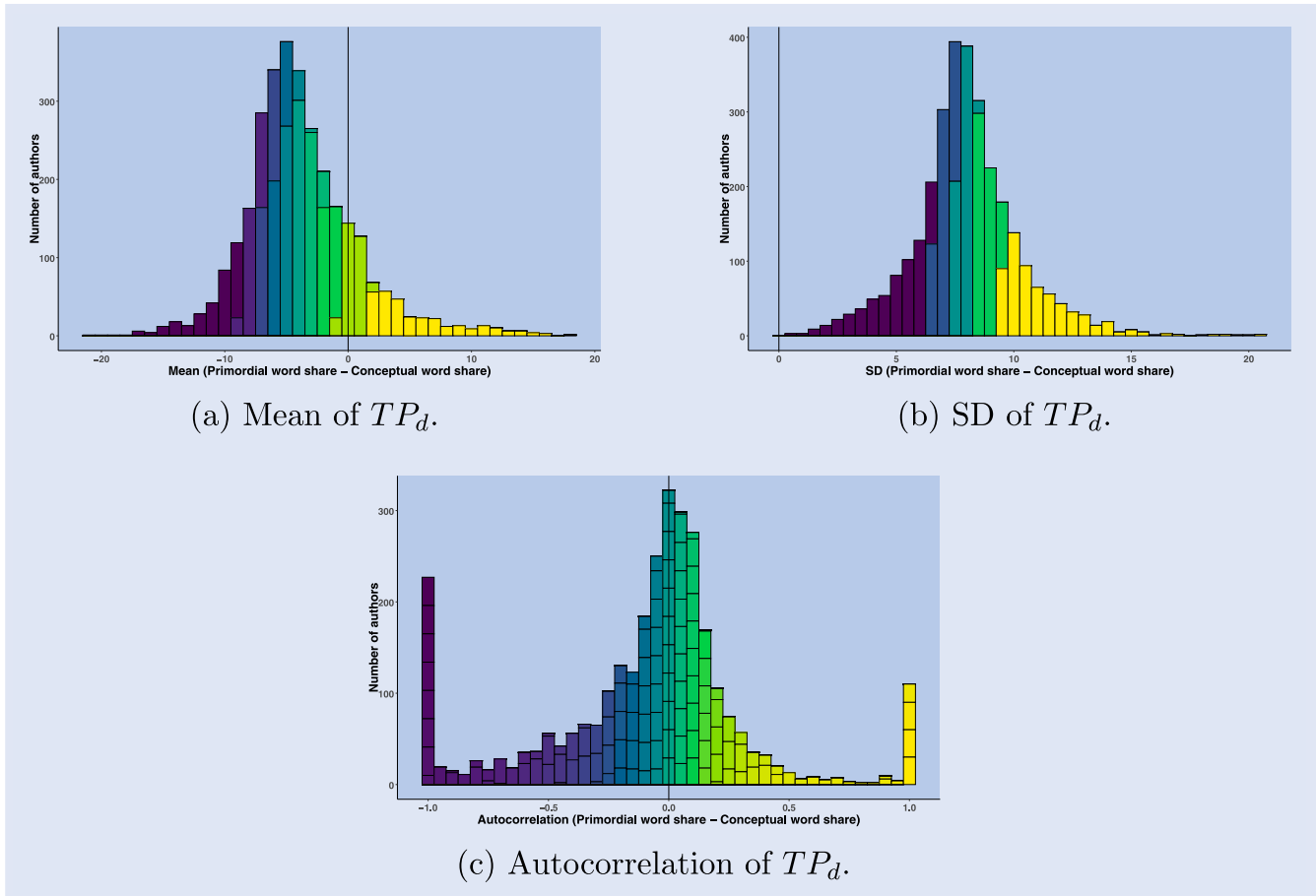


Figure 8. Distribution of the primordial - conceptual word share difference across authors. A total amount of 3062 authors and 74 articles per author on average. (a) Mean of TP_d . (b) SD of TP_d and (c) Autocorrelation of TP_d .

In equation (8), $\Delta_h Y_{i,t}$ denotes the change in the probability of a company i being in an information-sensitive state from period t to $t + h$. This probability is provided by the Gaussian Mixture Model discussed in section 2.1. The primary explanatory variable, $A_{k,t}$, represents the daily unexpected attention to topic k on day t . We quantify unexpected attention to a specific topic as the difference between the actual topic attention T_t^k and the predicted daily aggregate share $E(T_t^k | \xi_{t-1})$ of that topic in all articles from period t , given the prior news ξ_{t-1} .

$$A_t^k = T_t^k - E(T_t^k | \xi_{t-1}). \quad (9)$$

The coefficients β_k^h measure the impact of unexpected attention to news topic k on the change in the likelihood of being in an information-sensitive state across different horizons, while controlling for other news surprises.

We are also interested in identifying topics that serve as ‘systemic’ triggers. Their effect might be magnified when a company is specifically mentioned in the news. To address this in our analysis, we introduce the daily topic frequency $F_{i,t,k}$ among articles directly referencing company i on day t .

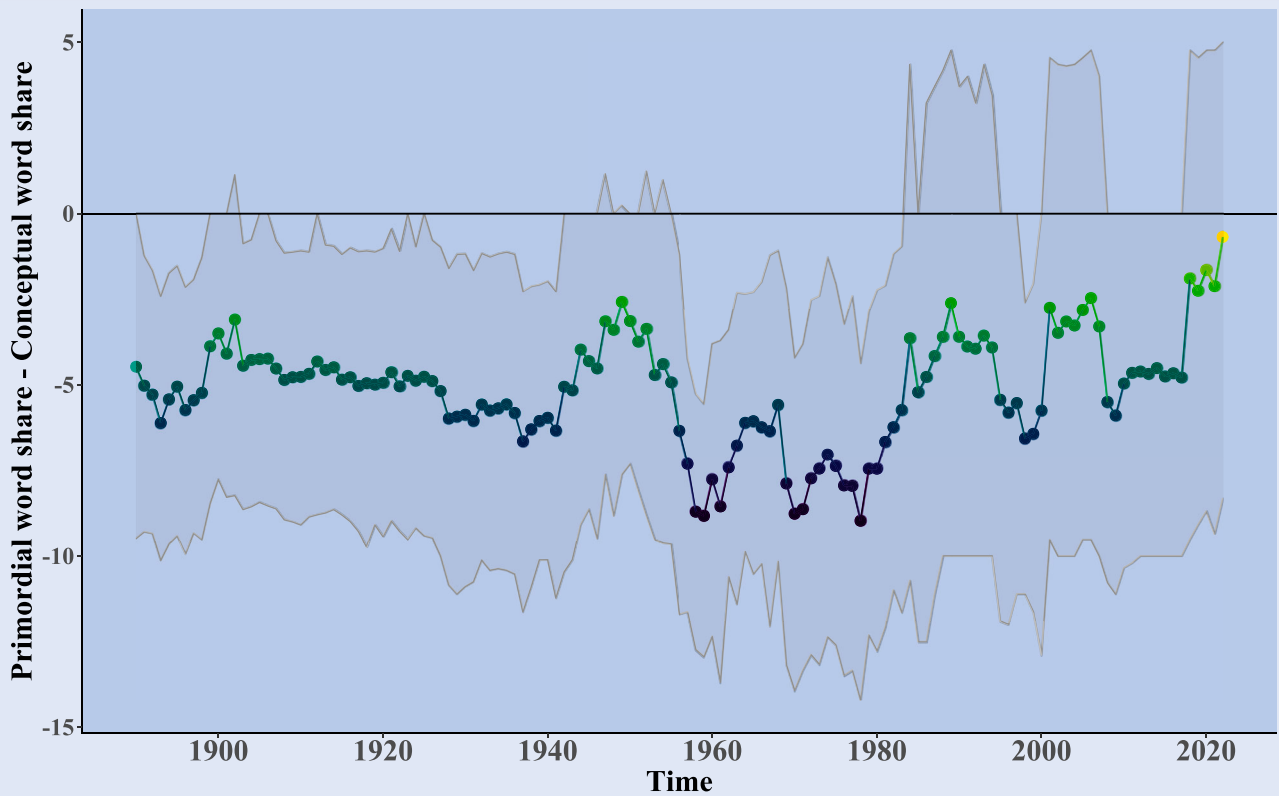
$$F_{i,t,k} = \frac{1}{N_{t,i}} \sum_{d=1}^{N_{t,i}} \theta_{d,k}. \quad (10)$$

This acts both as an individual explanatory variable and in conjunction with the unexpected topic attention $A_{k,t}$ for

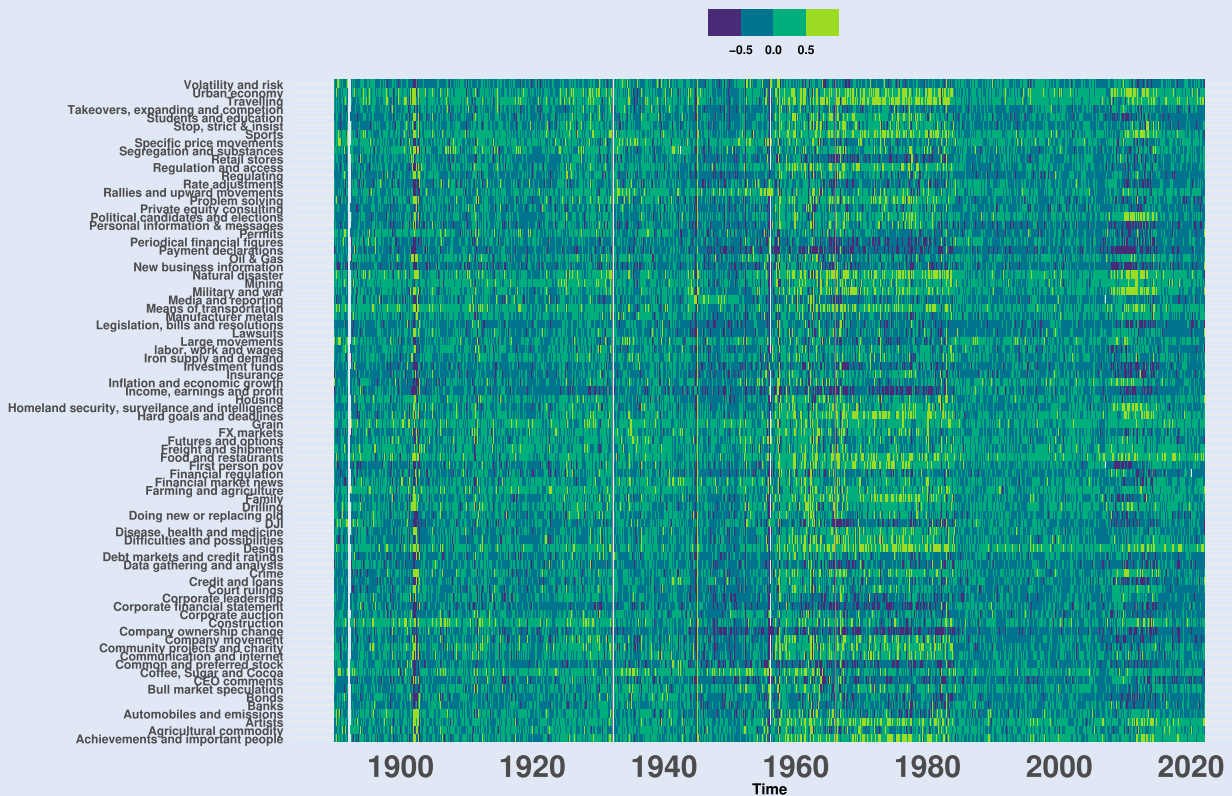
that day. The coefficient γ_k^h of the interaction term gauges the potential amplifying effect of direct company mentions on the probability of information sensitivity, resulting from unexpected topic attention in the news from h days prior.

The likelihood of a company entering an information-sensitive state can be influenced by a myriad of factors beyond the immediate, unexpected attention to news topics. Throughout our analysis, we account for a broad range of both aggregate level factors, denoted by Z_t , and company-level factors, denoted by $X_{i,t}$. These controls encompass variables that describe the general economic environment, such as past quarters’ GDP growth of various regions and bodies (e.g. US, OECD, OECD Europe, G-20, G-7). We also consider past financial market movements, specific economic sector returns, and company performance metrics, among others.† We’ve included company fixed effects, α_i , to account for

† These controls include variables describing past financial market movements (previous day, last 7 days and past 3 months) such as stock returns (SP500), market over/undervaluation (SP500 Shiller CAPE), uncertainty (SP500 volatility, VIX, Baker *et al.* (2016) GEPU and PUI indexes), returns on different economic sectors (Real Estate, Financial, Industrial, Energy, Utilities, Europe, Banks, Materials, Pharmaceuticals, Metals & Mining, Technology Hardware, Storage & Peripherals, Electronic Equipment, Software, Transportation) and also company level performance (stock return of past week and month), uncertainty (stock price volatility of past month and past six months) and current firm specific probability of an information sensitive state.



(a) Daily thinking process leaning across time. The figure plots the 25th and 75th (shaded area) quantiles and the average TP_t for each year in the period 1890–2022.



(b) Monthly correlation of topics and primordial - conceptual word share difference across time and topics.

Figure 9. Primordial-conceptual word share difference across time. (a) Daily thinking process leaning across time. The figure plots the 25th and 75th (shaded area) quantiles and the average TP_t for each year in the period 1890–2022 and (b) Monthly correlation of topics and primordial - conceptual word share difference across time and topics.

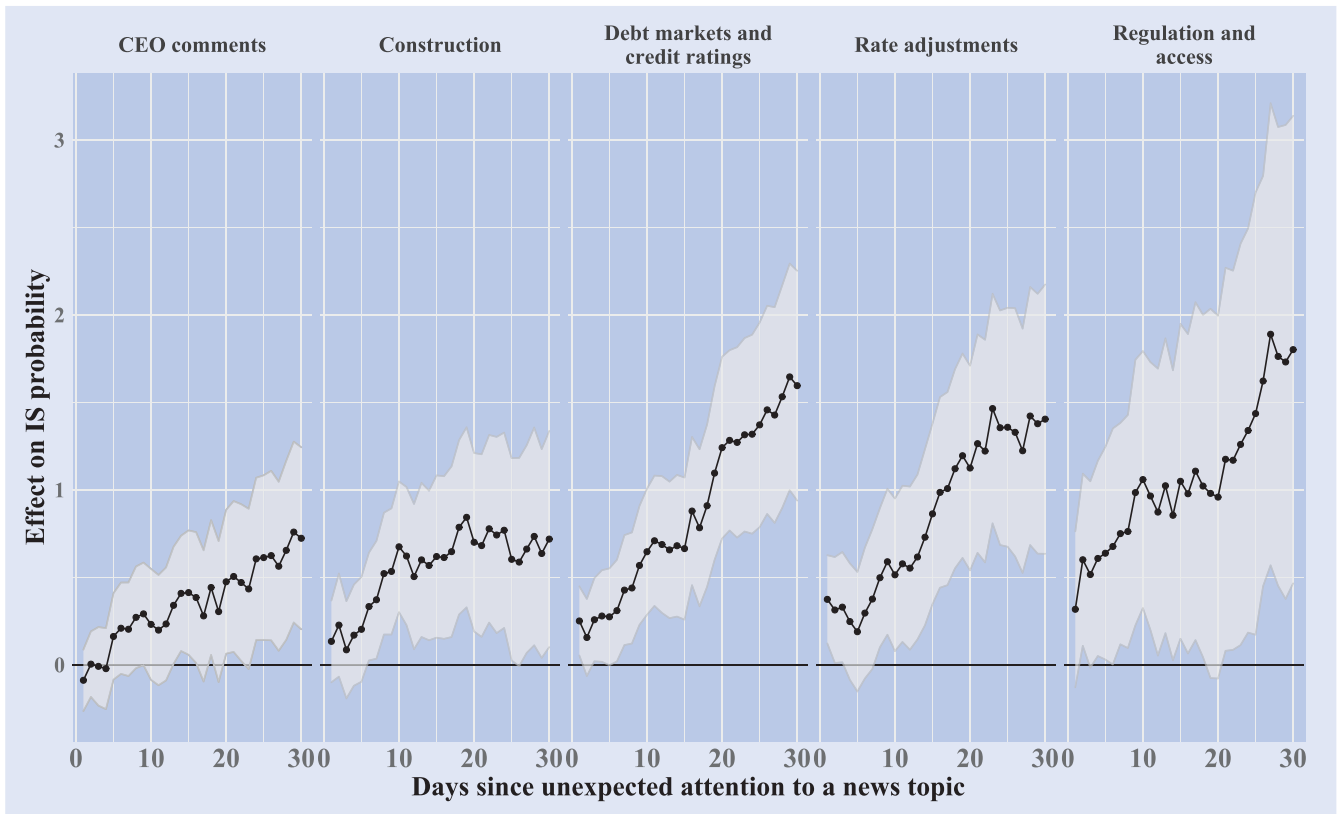


Figure 10. Narrative triggers of information sensitivity. The figure plots the β_k^h coefficients of equation (8) with 95% confidence intervals for topics with a positive and significant last period coefficients and at least 15 coefficients that are statistically significant at a 5% level between 1–30 day horizons. Statistical significance is calculated with standard errors clustered at the day and company level.

any firm-specific, time-invariant factors that might influence information sensitivity. Our analysis clusters standard errors at both the daily and company levels. The dataset underpinning our estimations consists of 884 721 day-company observations for 314 firms, spanning December 18, 2007, to February 4, 2022.

Figure 10 highlights the primary findings, showcasing the local projection coefficients, β_k^h , for topics that meet certain significance criteria.[†] Our approach prioritizes the most robust triggers over those that might have sporadic significant coefficients within the 1–30 day horizon. Notably, five topics — ‘CEO comments,’ ‘construction,’ ‘debt markets and credit ratings,’ ‘rate adjustments,’ and ‘regulation and access’ — stood out as logical, potential drivers of information sensitivity even before our analysis. Their prominence in our results further validates our methodology. The unexpected attention to these topics seems to trigger a gradual increase in the probability of becoming information sensitive in the upcoming days. This increase seems to start taking place around 5 days after the publication of the news after which the increase slows down or stops in around 10–20 days. Only the shock to the attention of ‘CEO comments’ topic increases gradually all the way up to 30 days after the news. A one percentage point positive deviation from the expected attention to a topic increases the probability of information sensitivity for a company around 0.5 to 2.0 percentage points. For instance, if ‘regulation and access’ topic would suddenly have a 10%

share in the news today in a situation where it was expected to have a share of 0%, there would be a 5 to 20 percentage point increase in the probability of companies being information sensitive. It’s worth noting that these are all systemic triggers; the actual company need not be the subject of such news articles. Figure 11 delves deeper into the ‘debt markets and credit ratings’ topic. It reveals that the coefficient γ_k^h is significant and positive solely for this topic, indicating a heightened sensitivity for companies directly mentioned in related news articles. This aligns with our expectations: news about credit rating downgrades has profound implications for a company’s information sensitivity. This is particularly true for firms whose ratings are directly impacted or speculated upon in such articles.

5.2. Economic performance and uncertainty

Following our empirical analysis of unexpected news attention’s impact on information sensitivity, we now examine whether these triggers operate uniformly across various economic or firm-specific situations, or if their effects are state-dependent. We enhance our panel local projection regression model, denoted by equation (8), to include interaction terms. Specifically, our terms of interest ($A_{k,t}$, $F_{i,t,k}$, and $A_{k,t}F_{i,t,k}$) are interacted separately with both a state-indicating dummy variable $S_{i,t}$ and its complement, $(1 - S_{i,t})$.

To clarify the nature of these states, we examine four separate specifications for $S_{i,t}$. Each specifies the state based on different metrics: previous quarter’s GDP growth, the daily

[†] Topics with positive and significant coefficient (5% level) at least for half of the days and for the furthest 30 day horizon.

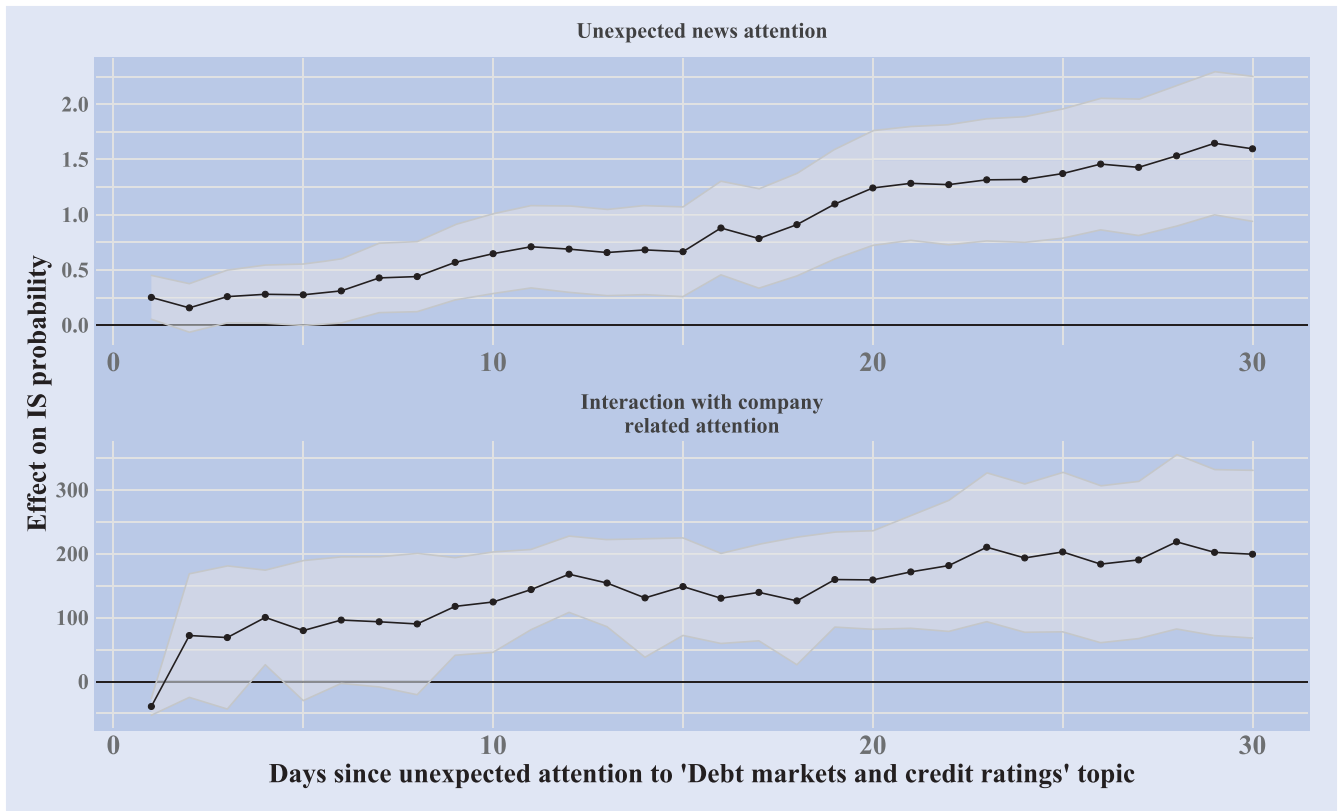


Figure 11. Individual company attention and narrative triggers of information sensitivity. The figure plots the β_k^h and the γ_k^h coefficients of equation (8) with 95% confidence intervals for topics with a positive and significant last period coefficients and at least 15 coefficients that are statistically significant at a 5% level between 1–30 day horizons. Statistical significance is calculated with standard errors clustered at the day and company level.

VIX index value, the firm's 30-day stock price volatility, and the firm's stock return from the previous week. The threshold for determining the state's strength is the sample median. A state is considered 'strong' or 'weak' based on whether the specific metric lies above or below this median.

Figure 12 displays the triggers that emerge as significant in at least one of these eight defined states, using a methodology consistent with our previous analysis. Notably, during economically weaker periods, these triggers are generally more potent. For instance, the 'CEO comments' topic becomes a trigger of information sensitivity predominantly during times of low growth or heightened uncertainty. Similarly, the 'debt markets and credit ratings' topic consistently emerges as a significant trigger, but its impact is markedly amplified during periods of economic downturn or increased uncertainty. The firm-level states (past week stock return and past months stock price volatility) seem to have a similar separating pattern for this topic, but the difference is not statistically significant as it is for the general economic states. The exactly same conclusions can be made for the triggering properties of the 'rate adjustments' topic.

Figure 13 offers insights into topics that either show no effect or diverge from our expectations. While some topics, such as 'urban economy', 'design', and 'food and restaurants', predictably display no significant reaction across any economic state, others were surprising. One might anticipate topics like 'lawsuits', 'court rulings', and 'new business information' to significantly influence information sensitivity

under certain conditions. However, the absence of significant effects underscores that genuine systemic triggers are indeed a rarity, confined to a select few topics.

5.3. Journalist dependent language

'What did eventually calm the European money markets? Governor Draghi's statement 'we will do whatever it takes – and you better believe it is enough.' This is as opaque a statement as one can have. There were no specifics on how calm would be reestablished, but the lack of specific information is, in the logic presented here, a key element in the effectiveness of the message. So was the knowledge that Germany stood behind the message – an implicit guarantee that told the markets that there would be enough collateral. A detailed, transparent plan to get out of the crisis, including rescue funds, which were already there, might have invited differences in opinion instead of convergence in views.'

–Holmstrom (2015)

Triggers of information sensitivity may not solely depend on specific topics, such as large sales or profitability movements, but also on the combination of the topic discussed and its presentation and perception by economic agents. For instance, if investors read a statement from a company's CEO regarding the firm's future plans during an economic downturn specific to that company (e.g. Nokia's strategy when

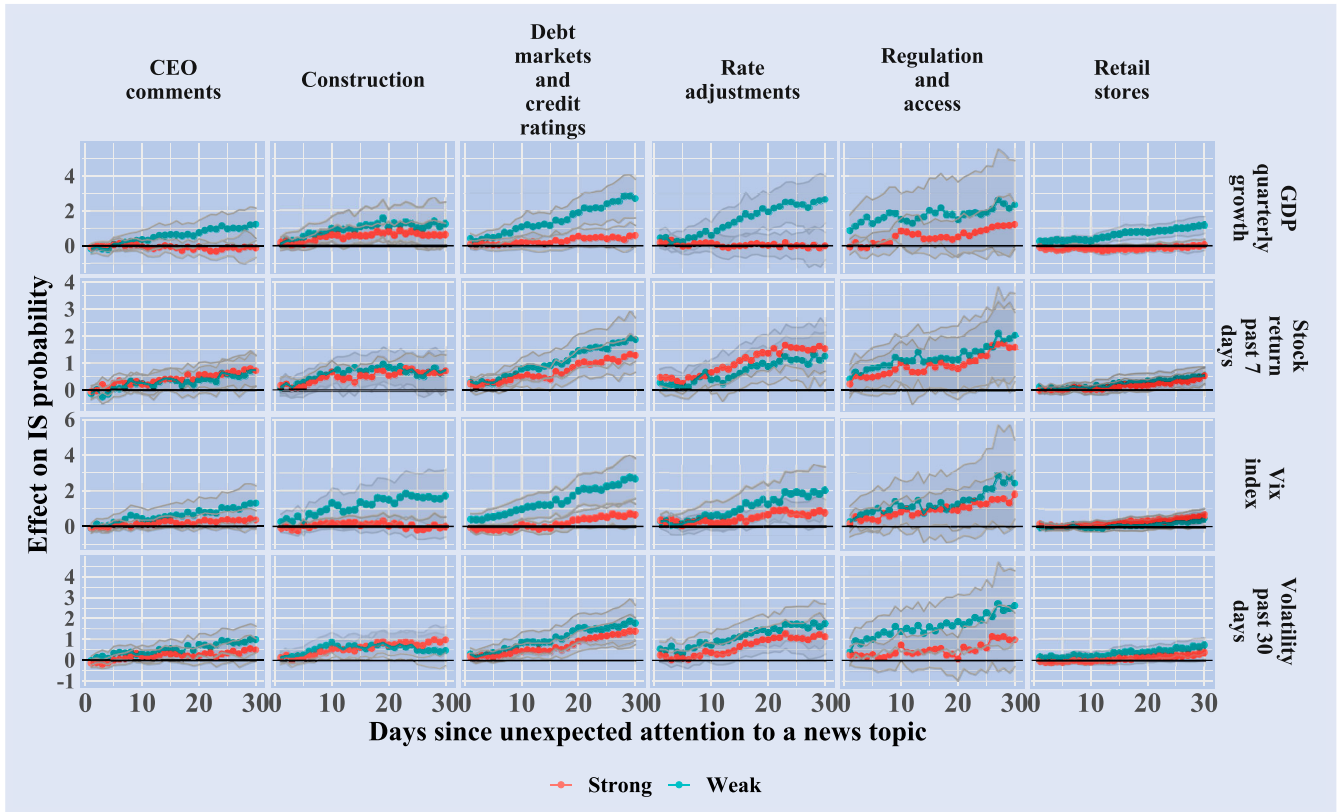


Figure 12. Narrative triggers of information sensitivity and different economic states. The figure plots the β_k^h coefficients of equation (8) with 95% confidence intervals for topics with positive and significant last period coefficient and at least 15 coefficients in total that are positive and statistically significant at a 5% level between 1–30 day horizons in at least one state. Statistical significance is calculated with standard errors clustered at the day and company level. The strong (red) coefficients refer to coefficients in the economically strong state and the weak (blue) refer to coefficients in the economically weak state.

Android and iPhone were rising in market share), the language chosen by the CEO, be it concrete or opaque and visionary, could be pivotal. Since information insensitivity and the operation of debt markets hinge on opaqueness, the language used can profoundly influence agents’ beliefs and the underlying fundamentals. A 2% decline in sales might be perceived differently depending on whether it’s characterized as ‘a rather modest decrease’ or ‘a never-before-seen drop.’ Similarly, news about a supply shortage in phone manufacturing materials could instigate information acquisition when described as a severe shortage lacking precise details. However, debt associated with mobile phone manufacturers might stay information-sensitive if the shortage is portrayed in accurate and relatively neutral terms.

To consider the potential influence of language on triggers, we estimate the local projection model in equation (8), incorporating data on the thinking-process-related language utilized in the articles. As delineated in the descriptive details from section 4, the thinking process and language evident in news articles show significant variance among journalists in our dataset. While most authors display a primary-conceptual thinking process language continuum, sizable groups (comprising hundreds of authors) lean more towards either a conceptual or primary thinking process language throughout their careers. We integrate this element into our analysis to ascertain if the thinking process and language chosen by journalists impact the emergence of certain topics as triggers of information sensitivity.

Our analysis unfolds as follows: We integrate the primary variables of interest with a language share variable, L_t , representing the average share of either primary or conceptual-related words in the news on day t . Formally stated:

$$\begin{aligned} \Delta_h Y_{i,t} = & \alpha_i^h + \beta_k^h \sum_{k=1}^{80} A_{k,t} + \beta_{l,k}^h \sum_{k=1}^{80} L_t A_{k,t} + \omega_k^h \sum_{k=1}^{80} F_{i,t,k} \\ & + \omega_{l,k}^h \sum_{k=1}^{80} L_t F_{i,t,k} + \gamma_k^h \sum_{k=1}^{80} A_{t,k} F_{i,t,k} \\ & + \gamma_{l,k}^h \sum_{k=1}^{80} L_t A_{t,k} F_{i,t,k} + \eta^h Z_t + \zeta^h X_{i,t} \\ & + \epsilon_t \quad \text{for } h = 1, \dots, 30. \end{aligned} \tag{11}$$

The $\beta_{l,k}^h$ coefficients related to the interaction terms $\sum_{k=1}^{80} L_t A_{k,t}$ reveal whether the triggering effect of unexpected attention to a topic is amplified with the increase in specific type of thinking process related language.

Figure 14 illustrates these coefficients for the topics that are significant triggers of information sensitivity and for which the language interaction coefficients were statistically significant (at the 5% level) for at least 15 periods and for the final 30-day period post-news publication. A discernible diminishing effect emerges from more considerable shares of primary thinking process language on the triggering effect of unexpected attention to ‘debt markets and credit ratings’ and

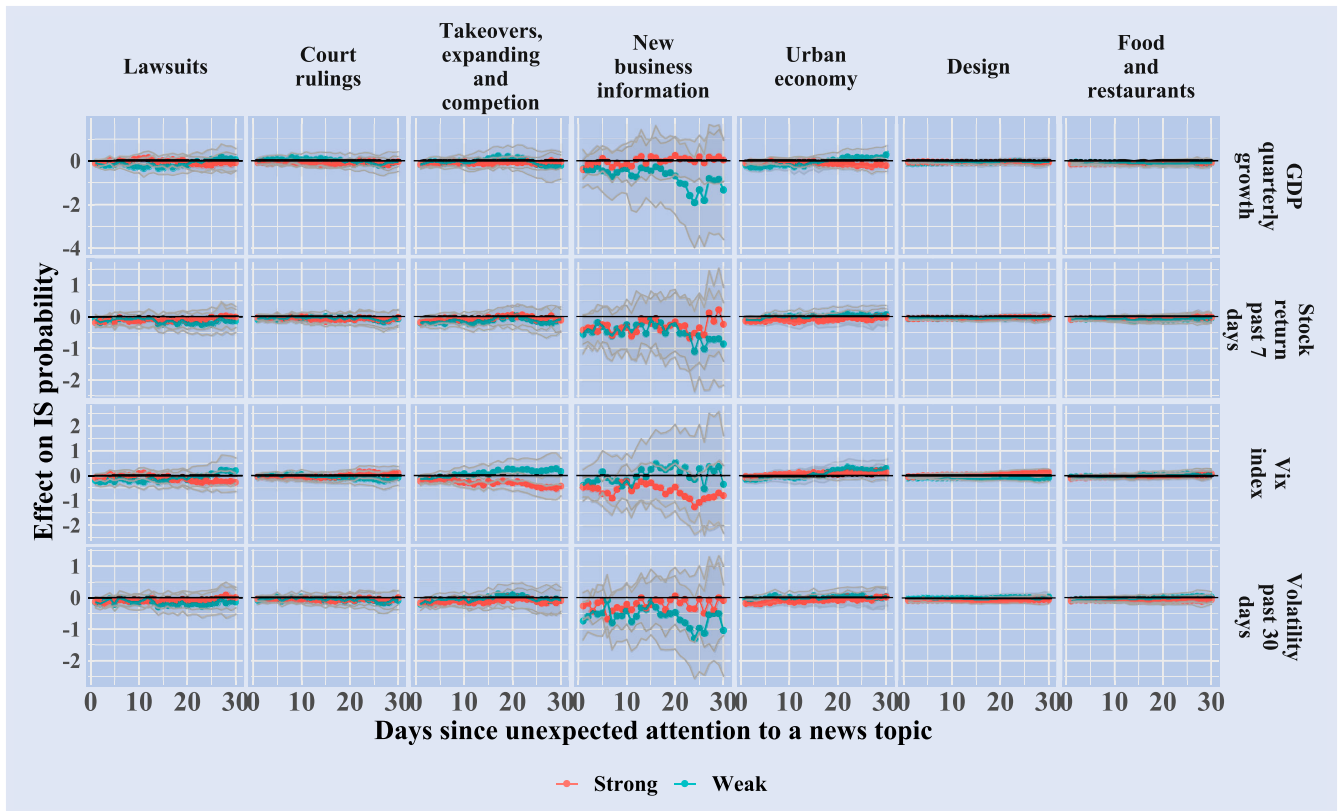


Figure 13. Narrative triggers of information sensitivity and different economic states. The figure plots the β_k^h coefficients of equation (8) with 95% confidence intervals for selected group of topics for which the coefficients are not statistically significant at a 5% level between 1–30 day horizons. Statistical significance is calculated with standard errors clustered at the day and company level. The strong (red) coefficients refer to coefficients in the economically strong state and the weak (blue) refer to coefficients in the economically weak state.

‘rate adjustments’ topics. This suggests that when journalists employ more creative, impulsive, aimless, or irrational language while discussing issues tied to credit ratings and interest rate adjustments, it becomes less probable that companies transition to an information-sensitive state. Conversely, in the event of a news attention spike related to ‘market speculation’, the otherwise escalating triggering effect weakens with increased usage of conceptual thinking process language. This implies that, for instance, during an abrupt influx of articles discussing the potential start of a bear market, the triggering effect diminishes when journalists use more rational, reality-oriented, problem-solving, logical, and narrowly-focused language.

5.4. Discussion

The empirical results of this section have highlighted three primary implications. First, certain topics serve as general systemic triggers of information sensitivity, such as ‘debt markets and credit ratings,’ ‘rate adjustments,’ ‘construction,’ ‘CEO comments,’ and ‘regulation and access.’ Notably, only for the ‘debt markets and credit ratings’ topic does mentioning a specific company significantly strengthen the systemic trigger for that individual company. These findings are consistent with our expectations concerning these information events, such as credit rating downgrades, interest rate hikes, new regulations,

or unforeseen CEO comments prompting economic agents to generally reassess companies’ outlooks.

Following a news attention shock, the triggering effect often rises gradually and tends to either decelerate or peak between 10 to 20 days post-publication. This lagged response suggests that economic agents might initially underreact to news shocks, with certain triggers exacerbating this phenomenon. As a shift to information sensitivity happens when economic agents choose to gather information about a firm based on their current knowledge set, this delayed reaction signifies an underreaction to news. Coibion and Gorodnichenko. (2015) have shown, through survey data, that professional forecasters’ consensus also leans towards underreacting to aggregate news. While our measure of unexpected attention to a news topic encapsulates aggregate news, we don’t possess a direct measure of consensus beliefs. Nonetheless, our information sensitivity metric mirrors economic agents’ decisions shaped by motivations and data related to a corporation, thus reflecting shifts in aggregate beliefs. Our empirical findings denote that general systemic triggers of information sensitivity don’t produce immediate impacts (on the same day). Instead, they introduce a sense of doubt among economic agents that manifests after a lag.

Second, these triggers present varying behaviors depending on a company’s and/or the aggregate economy’s current state concerning uncertainty and performance. Several triggers are

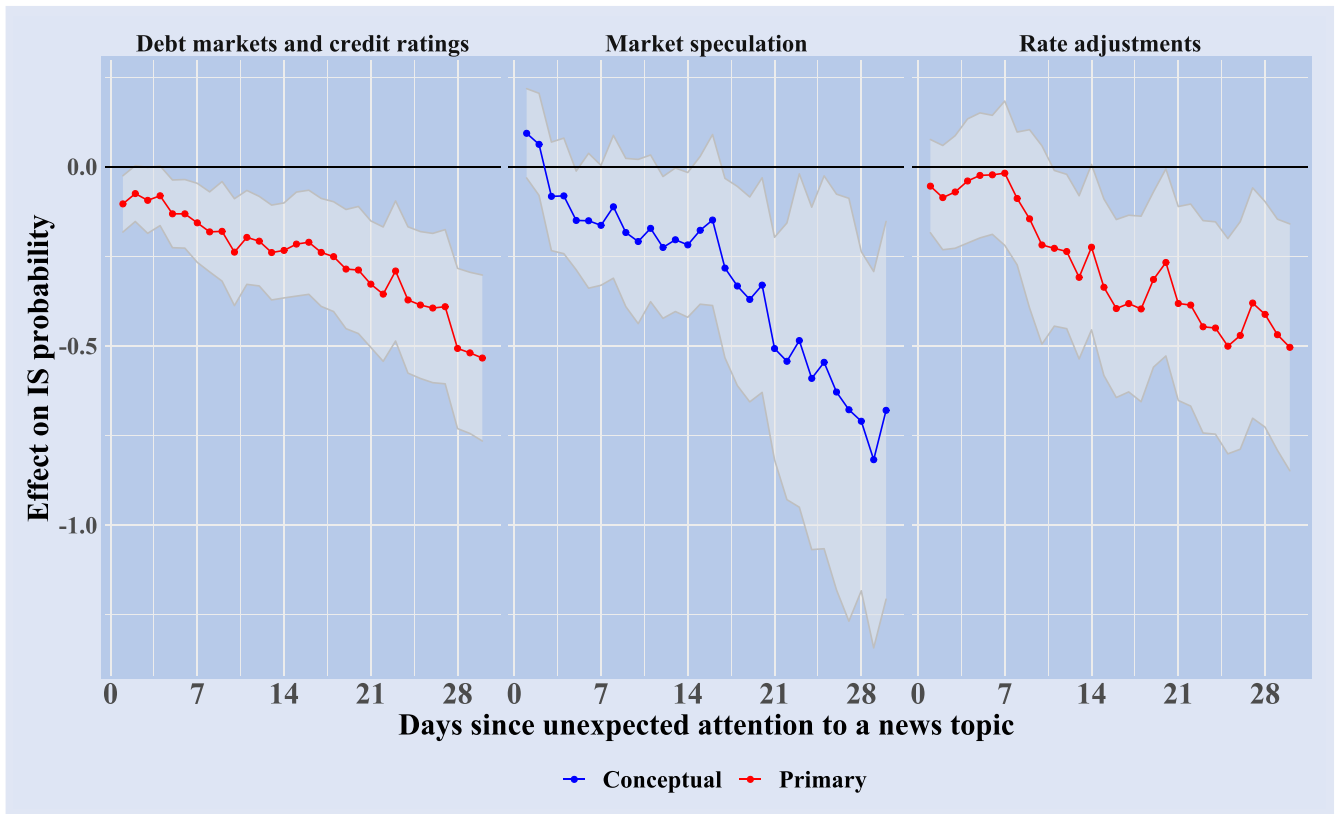


Figure 14. Journalists’ thinking process-related language and narrative triggers of information sensitivity. The figure plots the $\beta_{l,k}^h$ coefficients of equation (11) with 95% confidence intervals for statistically significant trigger topics where the last period, and at least 15 $\beta_{l,k}^h$ coefficients overall, are statistically significant at a 5% level between 1–30 day horizons. Statistical significance is computed with standard errors clustered at the day and company level.

more active during high uncertainty periods or when there’s poor performance, as indicated by low GDP growth or dismal stock returns. Some topics function as systemic triggers irrespective of the economic state, but their triggering effects are intensified during economic downturns.

Lastly, the thought process discernible from the language utilized in news articles notably affects whether a topic becomes a trigger of information sensitivity. Two notable systemic triggers—unexpected attention to ‘debt markets and credit ratings’ and ‘rate adjustments’—are diminished when described with more primary language. This pattern also appears in the ‘market speculation’ news (a systemic trigger of information sensitivity). However, the triggering effect reduces when more conceptual thinking process language characterizes the news.

If we assume news distribution among journalists is random and that our regressors capture the unpredictable component of attention to a news topic on any given day, these outcomes hint at a causal link between news topic attention, the employed language, and the prevailing information sensitivity in the economy. Given the affirmed efficacy of the regressive imagery dictionary in gauging a writer’s thought process across diverse scenarios and timeframes, it’s concerning that non-fundamental elements associated with news messengers can exert such a profound influence on the economic landscape. According to Freud (1938), an individual’s inclination towards the primary thinking process originates

from the urge to satisfy primary motives stemming from the id—the personality facet nurtured during early developmental years. Personal experiences and traumas from this period can subconsciously affect a journalist’s sentiments about a particular news topic. If the news content resonates with primary drives from the id, the reporter might address those drives by using more primary process language in their coverage.

These insights emphasize the interplay between news topics, economic scenarios, and language in shaping information sensitivity. Differentiating between writers who typically produce articles with characteristics ranging from irrational to rational, non-reality-based to reality-oriented, illogical to logical, impulsive to thoughtful, sensationalist to neutral, and aimless to purposeful, is vital. Given that information insensitivity is intrinsic to debt markets, ensuring their smooth operation, the indirect implications of these findings are concerning. Specifically, the effect of underlying events, possibly influenced by language variations due to writer-specific psychological nuances, might pose threats to financial stability and the broader economy.† When such journalists cover news topics frequently recognized as information sensitivity triggers, the distressing information event is missing.

† The possible implications for financial stability are deduced from the theoretical results of Dang *et al.* (2018, 2020).

6. Conclusion

In conclusion, we have provided insights into the triggers of information sensitivity in debt markets, the role of economic states, journalist language and thinking processes, and the dynamics of news content. By employing a comprehensive approach that combines quantitative analysis, machine learning techniques, and natural language processing, we have provided new insights on the relationship between news articles, information sensitivity, and journalists' thinking processes.

We begin by measuring the daily information sensitivity states of 576 financial and non-financial companies. Using machine learning methods with daily Credit Default Swap (CDS) spreads and Google search trends, we categorize each company-day observation into distinct information sensitivity states. This measurement approach captures the dynamics of information sensitivity for individual firms.

To identify the latent topics in news articles, we employ the Correlated Topic Model (CTM). This allows us to uncover underlying themes and patterns in news coverage. We find that news articles span a wide range of topics, including but not limited to economic indicators, geopolitical events, policy changes, and corporate developments.

To create the series of unexpected attention to news topics, we utilize a separate machine learning procedure that builds upon the output of the CTM. This procedure analyzes the daily prevalence of approximately 80 topics identified by the CTM. It identifies the portion of the daily frequency of each topic that could not be predicted by a machine learning model using past frequencies of all topics. This measure captures the unexpected attention given to specific news topics, indicating deviations from the predicted patterns. Examples of unexpected attention to news topics include events such as the global financial crisis, reporting on the Trump presidency, the COVID-19 outbreak, the start of the war in Ukraine, and the recent surprising burst in inflation. These examples serve to validate our methodology and demonstrate its ability to capture and quantify unexpected news events.

To analyze the triggers of information sensitivity, we delve into the language and thinking processes employed by journalists in news articles. Recognizing that news content is not solely determined by objective information, we explore how the writing style, creativity, and language choices of journalists can influence the signals received by economic agents. Drawing on psychological literature, we examine the connection between a journalist's personality and their writing style. We find evidence supporting the presence of primary and secondary thinking processes in journalists. The primary thinking process is associated with creativity, impulsiveness, and irrationality, while the secondary thinking process is rational, reality-oriented, and problem-solving. To measure the thinking processes of journalists, we utilize the regressive imagery dictionary developed by Martindale (1975), which distinguishes between words associated with primary and secondary thinking. Our analysis reveals that journalists exhibit varying degrees of preference for either the primary or conceptual thinking process in their writing.

Having explored journalist language and thinking processes, we then turn to the identification of triggers for information sensitivity in debt markets. Leveraging information

on aggregate and idiosyncratic uncertainty and economic performance, the measured differences in journalists language, and the series of unexpected attention to news topics, we conduct local projection analysis to investigate how specific news topics influence the probability of a company becoming information-sensitive. Our findings reveal that surprise attention to certain topics can act as systemic triggers of information sensitivity in the economy. We observe a lag of days between the occurrence of unexpected news attention and its effect on information sensitivity. Furthermore, the state of the economy or the firm, the language used by journalists, reflecting their thinking processes, plays a decisive role in determining whether a news topic acts as a trigger of information sensitivity.

The insights gained from this research have implications for policymakers, market participants, and researchers. By recognizing the influence of journalist language and thinking processes on information sensitivity triggers, we can improve risk assessment, enhance market surveillance, and gain a better understanding of the factors driving financial stability.

Acknowledgements

I thank seminar participants at the University of Turku, 2023 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics, and the empirical macro-finance session at the 2023 annual meeting of the EEA-ESEM (Barcelona) for their useful comments and discussions.

Disclosure statement

I received financial support from the Emil Aaltonen Foundation and Yrjö Jahnsson Foundation. Both foundations are funding general scientific research, without any particular financial, ideological, or political stake or, for example, providing additional accesses to (proprietary) data. I worked in the research unit of the Bank of Finland from 9/2017 to 8/2018. However, this specific research was not done or planned during my time in the bank. No other interest parties to declare.

Funding

I gratefully acknowledge the research assistance provided by Eetu Laakso and Robert Pylkkänen, as well as the financial support provided by the Emil Aaltonen Foundation and Yrjö Jahnsson Foundation.

Data availability statement

The paper uses text data from ProQuest historical newspapers. The actual text data is proprietary and it cannot be shared publicly. The data collection procedure is documented in

detail in the paper. The website for ProQuest Historical newspapers is <https://about.proquest.com/en/products-services/pq-hist-news/>. One needs to have a payed subscription to the database to get access to the actual text data. In addition, the CDS spread data is proprietary and it cannot be shared publicly. It has been collected from Refinitiv Datastream and one needs to have a payed subscription to the database to get access to the CDS data.

ORCID

Kim Ristolainen  <http://orcid.org/0000-0003-0470-1392>

References

- Baghai, R.P., Giannetti, M. and Jäger, I., Liability structure and risk taking: Evidence from the money market fund industry. *J. Financ. Quant. Anal.*, 2022, **57**(5), 1771–1804.
- Baker, S.R., Bloom, N. and Davis, S.J., Measuring economic policy uncertainty. *Q. J. Econ.*, 2016, **131**(4), 1593–1636.
- Behr, A., Assessing the stability of Gaussian mixture models for monthly returns of the S&P 500 index. *Appl. Financ. Econ. Lett.*, 2007, **3**(4), 215–220.
- Bianchi, F., Ludvigson, S.C. and Ma, S., Belief distortions and macroeconomic fluctuations. *Am. Econ. Rev.*, 2022, **112**(7), 2269–2315.
- Blei, D.M. and Lafferty, J.D., Correlated topic models. In *NIPS'05*, pp. 147–154, 2005 (MIT Press: Cambridge, MA, USA).
- Blei, D.M., Ng, A.Y. and Jordan, M.J., Latent dirichlet allocation. *J. Mach. Learn. Res.*, 2003, **3**, 993–1022.
- Brancati, E. and Macchiavelli, M., The information sensitivity of debt in good and bad times. *J. Financ. Econ.*, 2019, **133**(1), 99–112.
- Cipriani, M. and La Spada, G., Investors' appetite for money-like assets: The MMF industry after the 2014 regulatory reform. *J. Financ. Econ.*, 2021, **140**(1), 250–269.
- Coibion, O. and Gorodnichenko, Y., Information rigidity and the expectations formation process: A simple framework and new facts. *Am. Econ. Rev.*, 2015, **105**(8), 2644–2678.
- Curran, R., Moving the Market Banks Stocks Rally; Carnival Slips 12% Dow Industrials Rise 144.32, in First Two-Session Gain Since September, 2008. Copyright - (c) 2008 Dow Jones & Company, Inc. Reproduced with permission of copyright owner. Further reproduction or distribution is prohibited without permission (Last updated 20 November 2020).
- Dang, T.V., Gorton, G. and Holmström, B., Ignorance, debt and financial crises. Yale University Unpublished Working Paper, 2018.
- Dang, T.V., Gorton, G. and Holmström, B., The information view of financial crises. *Annu. Rev. Financ. Econ.*, 2020, **12**(1), 39–65.
- Dang, T.V., Gorton, G., Holmström, B. and Ordoñez, G., Banks as secret keepers. *Am. Econ. Rev.*, 2017, **107**(4), 1005–1029.
- Dougal, C., Engelberg, J., García, D. and Parsons, C.A., Journalists and the stock market. *Rev. Financ. Stud.*, 2012, **25**(3), 639–679.
- Freud, S., *The Interpretation of Dreams*, 1938 (Random House: New York).
- FT.com, Retailers punished as wall street sours on sector, 2015. *FT.com*. Copyright - Copyright The Financial Times Limited Nov 9, 2015 (Last updated 19 November 2020).
- Gallagher, E., Schmidt, L., Timmermann, A.G. and Wermers, R., Investor information acquisition and money market fund risk rebalancing during the 2011–2012 eurozone crisis. *Rev. Financ. Stud.*, 2020, **33**(4), 1445–1483.
- Glasserman, P. and Mamaysky, H., Does unusual news forecast market stress? *J. Financ. Quant. Anal.*, 2019, **54**(5), 1937–1974.
- Goldstein, K., *The Organism*, 1939 (Beacon: Boston).
- Gorton, G., Banking panics and business cycles. *Oxf. Econ. Pap.*, 1988, **40**(4), 751–781.
- Granger, C.W.J., The regressive imagery dictionary: A test of its concurrent validity in English, German, Latin, and Portuguese. *Lit. Linguist. Comput.*, 2011, **26**(1), 125–135.
- Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2009 (Springer).
- Holmstrom, B., Understanding the role of debt in the financial system. Bank for International Settlements BIS Working Papers 479, 2015.
- Jorda, O., Estimation and inference of impulse responses by local projections. *Am. Econ. Rev.*, 2005, **95**(1), 161–182.
- Kapner, S., Macy's seeks answers as sales slide—after resisting, CEO tries discount stores; stock tumbles on earnings, *Wall Street J.*, November 13, 2015, A.1. Name - Starboard Value LP; Macys Inc; Copyright - (c) 2015 Dow Jones & Company, Inc. Reproduced with permission of copyright owner. Further reproduction or distribution is prohibited without permission; People - Lundgren, Terry J (Last updated 13 September 2021).
- Katz, A.N., Creativity and the cerebral hemispheres. In *The Creativity Research Handbook*, edited by M. A. Runco, pp. 203–226, 1997 (Hampton: Cresskill, NJ).
- Kon, S.J., Models of stock returns—A comparison. *J. Finance*, 1984, **39**(1), 147–65.
- Kopcsó, K. and Láng, A., Uncontrolled thoughts in the dark? Effects of lighting conditions and fear of the dark on thinking processes. *Imagin. Cogn. Pers.*, 2019, **39**(1), 97–108.
- Lawton, C. and Efrati, A., Nokia's latest headache: Android, 2011. Copyright - (c) 2011 Dow Jones & Company, Inc. Reproduced with permission of copyright owner. Further reproduction or distribution is prohibited without permission (Last updated 22 September 2021).
- Malevergne, Y., Pisarenko, V. and Sornette, D., Empirical distributions of stock returns: Between the stretched exponential and the power law? *Quant. Finance*, 2005, **5**(4), 379–401.
- Martindale, C., *Romantic Progression: The Psychology of Literary History*, 1975 (Hemisphere: Washington, D.C.).
- Martindale, C., Syntactic and semantic correlates of verbal tics in Gilles de la Tourette's syndrome: A quantitative case study. *Brain Lang.*, 1977, **4**, 231–247.
- Martindale, C., Biological bases of creativity. In *Handbook of Creativity*, edited by R. J. Editor Sternberg, pp. 137–152, 1998 (Cambridge University Press).
- Martindale, C., Creativity, primordial cognition, and personality. *Pers. Individ. Differ.*, 2007, **43**(7), 1777–1785.
- Martindale, C., Covelto, E. and West, A., Primary process cognition and hemispheric asymmetry. *J. Genet. Psychol.*, 1986, **147**(1), 79–87.
- Martindale, C. and Dailey, A., Creativity, primary process cognition and personality. *Pers. Individ. Differ.*, 1996, **20**(4), 409–414.
- Mimno, D. and Lee, M., *Low-Dimensional Embeddings for Interpretable Anchor-Based Topic Inference*, pp. 1319–1328, 2014 (Association for Computational Linguistics, Doha, Qatar).
- Perignon, C., Thesmar, D. and Vuillemeys, G., Wholesale funding dry-ups. *J. Finance*, 2018, **73**(2), 575–617.
- Svensson, N., Archer, T. and Norlander, T., A Swedish version of the regressive imagery dictionary: Effects of alcohol and emotional enhancement on primary–secondary process relations. *Creat. Res. J.*, 2006, **18**(4), 459–470.
- Werner, H., *The Comparative Psychology of Mental Development*, 1948 (International University Press: New York).
- West, A., Martindale, C., Hines, D. and Roth, W.T., Marijuana-induced primary process thought in the TAT. *J. Pers. Assess.*, 1983, **47**(5), 466–467.
- Zou, H. and Hastie, T., Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, 2005, **67**(2), 301–320.

Appendix

A.3. Text preprocessing

A.1. Additional tables and figures

A.2. Text collection

- The data were collected from Proquest Historical Newspapers by using their TDM tool.
- The collection process was performed between 2022-03-28 and 2022-03-04.
- We collected all available titles, abstracts, page numbers, author names and dates of texts in the *Wall Street Journal* that we categorized as articles, features or news.
- The publication dates range from 1889-07-08 to 2022-02-05.
- After duplicated title-publication id pairs were removed, the corpus included 4 323 637 individual texts.

1. We identified empty titles, titles that were duplicates but still individual news (different publication id), and titles of irrelevant news types. Due to this, we removed the titles that included the following patterns: – No Title *OR* Dividends Rep *OR* ted *OR* Stocks Ex-Dividend—Stockholder Meeting Brief: *OR* Corrections & Amplifications: *OR* Corrections & Amplifications: *OR* TITLE BEGINS REVIEW *OR* TITLE BEGINS REVIEW & OUTLOOK (Editorial): *OR* TITLE BEGINS Business Brief: *OR* Theater: *OR* Dividend News: *OR* Sports *OR* Film *OR* Letters to the Editor: *OR* Co. TITLE ENDS *OR* Corp. TITLE ENDS *OR* Inc. TITLE ENDS *OR* Bookshelf: *OR* Opera: *OR* Gardening: *OR* Seeing Stars: *OR* Thinking Things Over: *OR* WORD BEGINS Art WORD ENDS *OR* TITLE BEGINS Books: *OR* Television: *OR* financial briefing book: *OR* reporter’s notebook.

Table A1. Topic labels and the predictability of the attention to topics 1–40.

Topic	Common words	MAE	Positive surprises %
Rate adjustments	Rate fix reduction percentage reduce adjust effect	0.186	60
Construction	Line construction facility terminal bridge construct connect	0.155	65
Hard goals and deadlines	Goal tough role tie deadline match hurdle	0.279	63
Freight and shipment	Import loading shipment coin gold bulk freight	0.114	76
Bonds	Certificate reorganization proceed obligation date bondholder coupon	0.186	58
Segregation and substances	Give pacific reading tobacco observe telegraph carbide	0.05	66
FX markets	Sterling mark check lira call guild belgian	0.134	66
Regulation and access	Free regulation commerce incentive barrier postal unlimited	0.065	71
Rallies and upward movements	Trend rubber narrow curb rally dealing early	0.168	53
Urban economy	Resident town city casino rural shelter gambling	0.28	65
Housing	Estate building property owner rent office tenant	0.295	57
Legislation, bills and resolutions	Tax adopt legislation bill resolution eliminate impose	0.312	63
Inflation and economic growth	Economy economic spending inflation economist boost euro	0.897	58
Credit and loans	Credit card loan lender borrower paper commercial	0.312	55
Difficulties and possibilities	Country difficulty circle possibility doubt circumstance prosperity	0.142	65
Corporate financial statement	Expect revenue cost gross result margin operating	0.301	51
Natural disaster	Fire disaster waste escape blame earthquake explosion	0.319	62
Oil & Gas	Natural gasoline gas energy pipeline gallon fuel	0.34	61
Financial market news	Small telephone clearing demand premium flat keen	0.061	58
Family	Family learn church mother son child lesson	0.357	71
Common and preferred stock	Stockholder stock common purchase prefer capitalization share	0.24	53
Court rulings	Decision ruling appeal answer legal question court	0.405	62
Permits	Permit road application railway grant permission necessary	0.104	68
Financial regulation	Financial regulator regulatory unit crisis collapse oversee	0.228	58
Stop, strict & insist	Rule mail implement stop strict insist editorial	0.097	65
Students and education	Student university op graduate education job academic	0.443	69
Doing new or replacing old	Take new step advantage symbol change helm	0.135	57
Lawsuits	bankruptcy Protection lawsuit filing file sue seek	0.291	59
Periodical financial figures	Increase figure year less compare period decrease	0.2	57
Disease, health and medicine	Drug study patient doctor disease cancer researcher	0.73	55
Sports	Team game player league pitch baseball football	0.73	59
Community projects and charity	Co complex community project found founder charity	0.147	66
Agricultural commodity	Wool textile sight cotton worth fertilizer south	0.069	63
Retail stores	Store retailer retail brand consumer shopper competition	0.353	59
Large movements	Rise drop jump climb lift surge index	0.442	51
Iron supply and demand	Iron pig inquiry concession basic foundry trade	0.07	60
Crime	Criminal prosecutor probe crime arrest jury guilty	0.8	60
Achievements and important people	Event honor revolution fortune moral self forget	0.334	66
Market speculation	Advance buying bull tendency selling speculative considerable	0.144	57
Labor, work and wages	Worker labor wage employ strike employment salary	0.178	67

Notes: The table presents the most common words for each topic calculated by the term frequency measure that takes into account the exclusivity and the frequency of each word. MAE refers to the Mean Absolute Error from the rolling window out-of-sample forecast made by the machine learning algorithm with information from past topic frequencies. The last column presents the percentage share of positive surprises (prediction errors) from all surprises.

Table A2. Topic labels and the predictability of the attention to topics 41–80.

Topic	Common words	MAE	Positive surprises %
Media and reporting	Press receive detail appear available comparison arrive	0.159	53
Specific price movements	Price level low high ounce pound depressed	0.24	57
Volatility and risk	Analyst emerge volatility investor risky cap volatile	0.418	51
Automobiles and emissions	Auto vehicle emission battery truck solar carbon	0.45	62
Design	Designer wall bedroom shirt clothe window style	1.339	54
Means of transportation	Air train airline flight plane driver passenger	0.436	61
Personal information & messages	Information personal employee message foundation define social	0.198	66
Artists	Music artist theater theatre novel song dance	1.215	65
Farming and agriculture	Farmer condition agricultural farm progress normal excellent	0.179	70
First person pov	Go get way many want one come	0.344	57
Regulating	Require law comply regulate state prohibit commissioner	0.207	66
Data gathering and analysis	Accord number datum analysis collect person release	0.155	62
Income, earnings and profit	Depreciation equal income equivalent net earn taxis	0.221	50
Communication and internet	Network ad search web advertising phone mobile	0.517	56
Corporate leadership	President board vice member chairman secretary senior	0.343	61
Private equity consulting	Partner strategy brokerage equity client top consulting	0.241	52
Debt markets and credit ratings	Term raise debt rating finance swap downgrade	0.216	58
Investment funds	Invest mutual investment fund pension manage asset	0.538	55
Coffee, Sugar and Cocoa	Point close bag unchanged sugar steady closing	0.099	55
Corporate auction	Bid municipal corporate auction offer bidder syndicate	0.219	58
Company ownership change	Plan acquire merger announce company ownership shareholder	0.288	55
Military and war	Military army troop regime rebel militant civilian	0.746	67
Travelling	Visit cross trip vacation rich bind obstacle	0.159	62
Payment declarations	Pay declare distribution payment declaration annual sum	0.146	60
Grain	Bushel wheat northwest grain winnipeg argentine visible	0.05	62
New business information	Business make understand important fact present enterprise	0.163	60
Political candidates and elections	Party election campaign presidential candidate voter debate	0.95	63
Takeovers, expanding and competition	Stake takeover venture expand customer competitor compete	0.311	51
Insurance	Care insurance familiar people insurer coverage matter	0.341	69
Drilling	Drill sand produce drilling spill gravity deep	0.14	64
Banks	Banking bank banker saving institution currency central	0.281	55
Food and restaurants	Food restaurant ice eat drink cup dog	0.886	56
Mining	Coal copper mine ore lake shipping vessel	0.134	66
Homeland security, surveillance and intelligence	Agency official administration citizen protect intelligence ministry	0.351	67
Dow Jones Industrial Average	Current value industrial average appreciation list secondary	0.103	69
Futures and options	Contract near commodity deliver dealer derivative position	0.1	53
Problem solving	Solution problem tool process experiment try rely	0.22	60
CEO comments	Chief executive address interview conference say comment	0.303	61
Company movement	Begin put set recent several bring similar	0.03	59
Manufacturer metals	Material steel manufacturer capacity manufacture scrap tin	0.108	59

Notes: The table presents the most common words for each topic calculated by the term frequency measure that takes into account the exclusivity and the frequency of each word. MAE refers to the Mean Absolute Error from the rolling window out-of-sample forecast made by the machine learning algorithm with information from past topic frequencies. The last column presents the percentage share of positive surprises (prediction errors) from all surprises.

- Extra whitespace was removed from texts, and all letters were changed to lowercase.
- If the abstract was not missing, then it was chosen as the text representing the article; otherwise, the title was chosen.
- Non-duplicate texts with at least 20 words were included.
- Python's Spacy library was utilized to parse the individual texts into individual parts of a sentence and identify the final list of words that we wanted to include.
- All words with the following entity categorization were removed: CARDINAL, DATE, EVENT, FAC, GPE, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL, ORG, PERCENT, PERSON, PRODUCT, QUANTITY, TIME and WORK_OF_ART.
- All words that had the following universal tag for parts-of-speech were chosen for inclusion: Adjectives (ADJ), Nouns (NOUN) and Verbs (VERB).
- The remaining words in each text were transformed into their lemma form.
- Lemma forms that were stopwords, included only one character, or included numbers or punctuation were removed.
- The lemmas that were among the lemmas with the highest 10 000 term frequency-inverse document frequency (tf-idf) values were included in the final corpus.
- Texts that had fewer than 10 words/lemmas after the cleaning process were removed from the final corpus.