Effects of Information Overload on Financial Markets: How Much Is Too Much?*

Alejandro Bernales   Marcela Valenzuela   Ilknur Zer

January 2024

Abstract

Motivated by cognitive theories showing that investors have a limited processing capacity, we study the effects of information overload on stock returns. We use textual analysis and machine learning tools applied to *The New York Times* since 1885 to measure information overload and structure our empirical analysis around a discrete-time learning model. We find that investors learn from news and their decision quality improves with information flow, but only up to a threshold. Excessive information load exhausts investors' processing capacity and deteriorates their decision accuracy, leading to a higher market risk premium. Information overload also has cross-sectional effects: stocks that are more difficult to value and require large learning efforts have larger expected returns.

Keywords: Limited attention, sentiment, predicting returns, behavioral biases, media content
JEL classification: G40, G41, G12, G14

*Alejandro Bernales is at Universidad de Chile (Facultad de Economía y Negocios, Departamento de Administración). Marcela Valenzuela is at the School of Management, Pontificia Universidad Católica de Chile. Ilknur Zer is at the Federal Reserve Board. We thank Theis Ingerslev Jensen, Charles Calomiris, Paul Tetlock, Thierry Foucault, Dong Lou, Ioanid Rosu, Svetlana Bryzgalova, Ana Elisa Pereira, Consuelo Silva, Felipe Aldunate, Carlos Canon, George Skiadopoulos, Pablo Neudorfer, Richard Holowczak, Louisa Chen, and seminar participants at the Federal Reserve Board, 5th Future of Financial Information Conference at HEC Paris, 7th Annual News and Finance Conference at Columbia Business School, 16th Behavioural Finance Working Group, and University of Chile. We also thank Maricel Vargas for her valuable assistance in data processing. Valenzuela acknowledges the support of Fondecyt Project No. 1231630 and the support of Instituto Milenio ICM IS130002. Bernales acknowledges the funds provided for the Fondecyt Project No. 1231660. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/R009724/1]. The views in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1 Introduction

The traditional asset pricing theory assumes that prices incorporate all available information instantly. However, in our daily lives, we have limited resources and are plagued with excess information. Thus, in reality, processing information requires sufficient resources such as time, budget, or attention from investors. As quoted by Nobel laureate Herbert Simon, "a wealth of information creates a poverty of attention". Given that investors have limited processing capacity (Kahneman, 1973; Johnston and Pashler, 1998), does information overload affect their investment decisions?

In this paper, we study the role of news media on stock market dynamics, while focusing on the load of information agents face. Using textual analysis and logistic regression machine learning tools, we scan the full content of *The New York Times* since 1885 and measure the information load as the total number of financial market articles published in a given day.[1] We use these novel data to study the effects of "excess" information (information overload) on excess market returns.

Why does information overload affect excess returns? As a roadmap for our empirical analysis, we introduce a discrete-time learning model. The model includes a single asset and a representative investor with a *limited* processing capacity, who learns from news and update her beliefs on the asset value according to Bayes' rule. Importantly, the model allows a nonlinear relationship between information load and the number of articles that the investor can actually process. The learning process improves with new information, increasing the precision of posterior beliefs. However, when the number of news exceeds a threshold, the investor becomes cognitively overloaded in line with the predictions of the early cognitive

[1] One can argue that excess information flow from mass and social media is one of the salient features of the modern information age only. Yet, the phenomenon is not confined to the modern world (Roetzel, 2019; Gleick, 2011; Blair, 2012). For instance, according to Blair (2012), even in the 13th century, information overload was present in the form of "the multitude of books, the shortness of time, and the slipperiness of memory."

scientists such as Schroder et al. (1967); Miller (1956); Simon and Newell (1971). When facing excessive information, the investor is overwhelmed and *not able to* analyze some useful information. Alternatively, a cognitively-overloaded investor may *choose not to* process new information.[2] Therefore, in our model, the additional information diminishes the investor's learning capacity and the precision of the posterior beliefs, implying that the estimated parameters of the true asset value are more likely to be wrong. Consequently, even with constant risk aversion, the risk-averse investor would require a higher risk premium to hold the asset in periods of information overload because of the increased estimation risk.

As a prelude to our empirical analysis, we show that information load (*InfLoad*)—the total number of financial markets-related articles published in a given month—is significantly correlated with various proxies of information and estimation risk, including dispersion in household expectations, bid-ask spreads, and analysts' forecast errors. It captures expected trends in the information flow, such as changes in editorial decisions, the introduction of the internet or other news outlets, and major economic events.

We then measure information overload (*InfOver*) as the number of articles above its one-standard deviation band using twelve months of moving windows as the threshold. The main findings are robust to choosing a different moving window size or using historical mean and Hamilton (2017) filter as thresholds. Importantly, by using a historical moving window threshold, we implicitly assume that investors can learn to digest more or less information as they become accustomed to it, altering agents' attention capacity, resources, and learning abilities.

We find that although receiving information reduces the market risk premium,

[2]In recent neuroscience literature, Reutskaja et al. (2018); Callicott et al. (1999); Jaeggi et al. (2007) use functional magnetic resonance imaging to study how human brain processes information. The images suggest an inverted V-shaped relationship between the brain activity and cognitive load, suggesting a turning point where processing new information outweighs the benefits.

when the number of news exceeds the threshold, i.e., when investors face information overload, they require a higher market risk premium. The relationship is economically meaningful: a one standard deviation increase in *InfOver* increases the monthly market risk premium by almost 60 basis points. *InfOver* significantly explains the market risk premium during pre- and post-World War II and pre- and post-mid 1990s, periods with significantly different information flow dynamics. Moreover, between the quantity and quality (the tone consistency) of the news, the quantity matters the most when predicting market returns.

Furthermore, we show that information overload does not affect all stock returns uniformly and instead, it has cross-sectional effects. Specifically, the impact of *InfOver* is stronger for smaller and more volatile stocks, and stocks with a higher share of retail investors and those not included in the S&P500 index, compared to their peers. These findings suggest that a plethora of information makes valuation even more difficult and exacerbate estimation risk for stock that require larger learning efforts. Our results are in line with Akbas et al. (2018), who show that information processing is more difficult for stocks with lower institutional ownership, and Hong and Sraer (2016); Baker and Wurgler (2006) who argue that small and riskier stocks are more difficult to arbitrage. Moreover, we find supportive evidence for the category learning model of Peng and Xiong (2006), in which cognitively constrained investors allocate their capacity to a certain group of stocks that require large learning efforts.

While documenting the impacts of information overload on stock returns, we face two identification challenges. First, our analysis assumes that the supply of news is exogenous. Nevertheless, it can change with macroeconomic conditions because demand for news is higher during turmoil periods. It could be macroeconomic conditions and uncertainty that drive the market risk premium, rather than information overload. Second, we use only the coverage in *The New York Times* to quantify information overload, and thus we do not consider the flow of information from other public resources, such as other news outlets, newspapers, or the internet. To alleviate such concerns, we conduct further analyses, which boost our confidence in our main finding that information overload increases excess stock returns in the next period.

First, we include various control variables to account for the macroeconomic and financial conditions in our specifications, including NBER recession dates, term spread, and the market sentiment (SENT) measure of Tetlock (2007) and Garcia (2013). In addition, we control for other news-based measures such as the Economic Policy Uncertainty index (EPU) of Baker et al. (2016), the Geopolitical Risk Index (GPR) of Caldara and Iacoviello (2022), and the News-implied Volatility Index (NVIX) of Manela and Moreira (2017). Furthermore, as robustness checks, we consider inflation rate, changes in the industrial production index, and market risk perception measured by the Duration of Low Risk (DLR) of Danielsson et al. (2023). We find that *InfOver* contains additional explanatory power on market returns beyond the standard predictors of returns, macroeconomic factors, and other news-based measures. We also verify the robustness of our findings when we exclude major stress periods, such as the Great Depression, and Global Financial Crisis, and stock market index crashes.

We also attempt to address the endogeneity concerns via two additional approaches. We start by employing two-stage least squares (2SLS) regressions by using alternative news (such as news on obituaries, fashion, and gaming) and generate instruments for *InfLoad* and *InfOver*. Here, our identification assumption is that these alternative news are unlikely to directly affect stock market returns and their supply is not driven by macroeconomic conditions but rather by editorial choices or the structure of the newspaper at a time. Furthermore, we test the effects of information overload on future stock market returns during exogenous events such as newspaper strikes or format changes. When an exogenous event reduces the supply of news (i.e., *InfOver* is lower), we find that the effect of *InfOver* on market returns is diminished. We conclude that our main findings continue to hold under either approach.

We then attempt to address the second identification challenge we face—that information overload is measured by relying only on the printed edition of a single newspaper. We are undoubtedly aware that *The New York Times* data do not comprehensively capture the entire flow of information. Nevertheless, even by relying only on the printed edition of a single newspaper, we document a robust relationship between information overload and stock returns. Thus, considering the additional sources would only exacerbate information overload and its effects on stock markets. Still, we verify the information overload–stock market relationship in a different setting by measuring information overload through the wave of earnings announcements made by the universe of U.S. publicly traded firms. We show that when many firms release earnings reports simultaneously in a day (investors face information overload), it becomes challenging for investors to focus on each announcement adequately so that the average contemporaneous stock price is lower in comparison to the days in which fewer firms are announcing.

This paper makes two important contributions. First, it builds upon an extensive body of literature from various disciplines, including neuroscience, organization science, accounting, and marketing that study the effects of information load on agents' decision quality (Eppler and Mengis, 2004; Edmunds and Morris, 2000; Grisé and Gallupe, 1999; Loughran and McDonald, 2014; Reutskaja et al., 2018; Lee, 2012). Here, we introduce an application of these principles to finance by studying the effects of information overload on stock market returns. Second, we construct a news-based historical index that measures the level of (excess) information agents face, which can arguably be used as a proxy for agents' limited attention as in DellaVigna and Pollet (2009); Hirshleifer et al. (2009); Da et al. (2011).

Overall, we conclude that information overload explains both time-series and cross-sectional variations in stock market returns, suggesting that prices do not incorporate all available information instantly as opposed to the conclusions of traditional asset pricing theories. It is worth emphasizing that our results do not necessarily imply a behavioral bias. Price reactions driven by information overload are different from the behavioral bias reaction of Tetlock (2007) and Garcia (2013), which argue that investment decisions and price changes reflect market sentiment. Similarly, we depart from the behavioral model of Hong and Stein (2003), in which investors have different opinions driven by overconfidence. Instead, we argue that information load affects returns through the constraints in investors' information processing capabilities, similar to the delay in the impounding of information into asset prices due to the complexity of information in Cohen and Lou (2012). It may also provide a potential explanation for the gradual information flow mechanism of Hong and Stein (1999).

This paper is organized as follows. In section 2, we introduce the construction of *InfLoad* along with descriptive analyses. Section 3 presents a discrete-time learning model, which we use as a guide for our empirical analysis. In section 4, we first describe our econometric methodology, introduce the variables used in the regressions, and then present the results. Finally, we conclude in Section 5.

2 Measuring information load

We construct the information load index in four steps. First, we scan the *full* content of daily *The New York Times* newspaper from January 1, 1885, to December 31, 2022. Data are obtained from the printed edition of *The New York Times*, ProQuest, TDM Studio. We obtain titles, keywords, and the lead paragraph of each article published.[3]

Our second task is to distinguish the financial and economic news ("business news") from other news (i.e., sports, weather, etc.). Post-1981, the name of the corresponding section of the article is provided, which enables us to identify the business news. For the pre-1981 period, we classify each article using the Logistic Regression machine learning tool.[4] Specifically, we train the tool on the articles that already have section names, so that it can learn how to classify the rest of the articles. We have ended up with a total of 2,214,296 business news articles.

Third, within the business news, we focus on news related to financial markets only, by adopting the word clouds of Calomiris and Mamaysky (2019). The authors employ the Louvain method (Blondel et al., 2008), which assigns salient words to mutually exclusive topic areas based on word co-occurrence and divides words into topic groups. They provide word clouds for the five topic groups: markets, governments, commodities, corporate governance and structure, and the extension of credit. We adopt the word clouds of the topic *markets*.

Clearly, an article can contain news on more than one topic (commodities, governments, etc.). Thus, following Calomiris and Mamaysky (2019), for a given article j published in day d, we assign a weight corresponding to the topic markets (Mkt):

$$w_{Mkt,j,d} = \frac{C_{Mkt,j,d}}{C_{j,d}}, \tag{1}$$

where $C_{Mkt,j,d}$ and $C_{j,d}$ are the number of words associated with topic markets and the total number of words appearing in the lead paragraph of article j in day d,

[3]Given the extensive data and the fact that the main message and tonality must be set out in the first paragraph as noted in the *The New York Times* writing practices, we obtain the lead paragraph as in Chan (2003) as opposed to the full article. See also `https://archive.nytimes.com/www.nytimes.com/learning/general/weblines/411.html`

[4]We apply Neural Networks, Gradient Method, and Logistic Regression algorithms. After training, testing, and validating algorithms, we conclude that the Logistic Regression has the best performance, with an accuracy of 92.5%, and thus we use the optimized parameters from the Logistic Regression model to classify the news.

respectively. Thus, for each article, and a day, we calculate the relative frequency of words that correspond to markets.

Finally, we define monthly information load as the total number of financial market articles published in a given month t:

$$InfLoad_t = \sum_{j,d} w_{Mkt,j,d}. \tag{2}$$

2.1 Descriptive analysis — information load

Figure 1 plots *InfLoad*, averaged across days in a given year since 1885. Broadly, the total number of printed articles related to financial markets follows an inverted-W shape: increasing early in the sample and decreasing post-1930s, followed by a spike in 1980s and decreasing then after. In our empirical analysis, we normalize the number of financial market news to account for this variation, which is likely driven by changes in the structure/length of the printed edition of *The New York Times*. Specifically, we consider the deviation of news from a historical trend when calculating information overload.

The number of financial market news was broadly unchanged until the early 1900s, followed by an increasing trend following the end of World War I. The 1920s, also the so-called "Roaring Twenties", was a period of exuberant economic and social growth in the United States. From 1920 to 1929, until the stock market crash, stocks more than quadrupled in value. Relatedly, the number of financial markets news also soared during that period and started to decrease then after. Moreover, by 1930, about 40 percent of the U.S. population owned a radio, using it as another means of news outlet.

Similarly, during World War II, although political and economic news coverage increased overall, we see a broadly decreasing trend for the financial markets news. We see another drop in 1962 for a short period because of the newspaper strike in New York City. During the late 1970s, the paper changed its column format, causing the number of articles to increase significantly. During mid 1980s, the first digital production of *The New York Times* has started and post-mid 1990s the usage of the internet and other news outlets such as Bloomberg or social media changed the use of newspapers. Yet upward trends are observed even in recent periods, such as during the 2008 Global Financial Crisis.

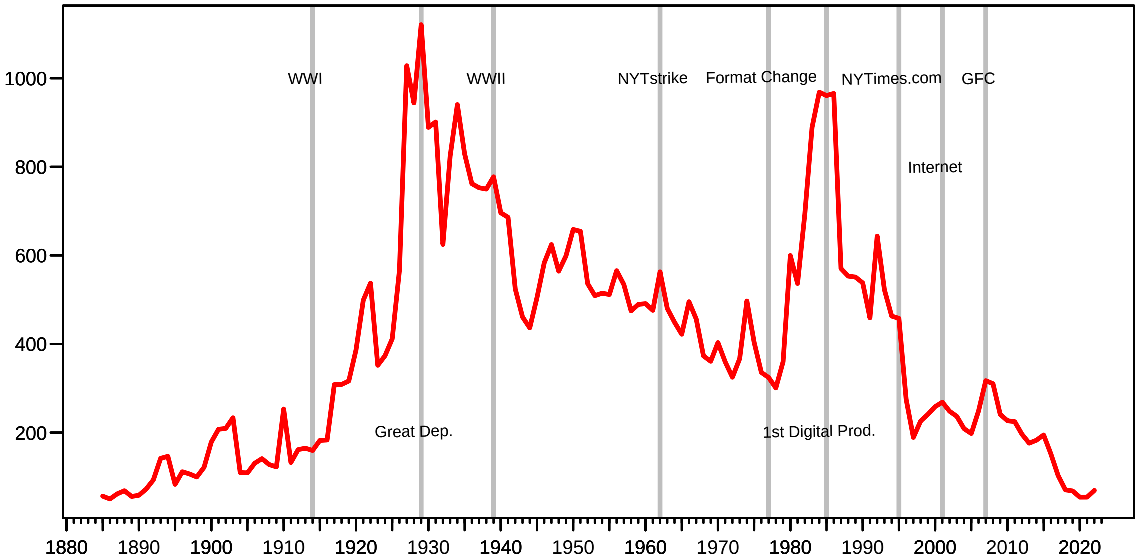


Figure 1: Information Load

The figure presents the annual averages of the total number of financial markets-related articles published in a given month. The sample period is 1885–2022. Data are obtained from the printed edition of *The New York Times*, ProQuest, TDM Studio. New York Times Historical Newspapers.

Furthermore, in Table 1, we present contemporaneous Pearson correlation coefficients of information load, with proxies for (1) information risk; (2) estimation risk; and (3) financial stress. First, a high information load should be associated with high information risk, as it compromises investors' ability to process information (see, for example, Bawden and Robinson, 2009; Muslu et al., 2015). To proxy the information risk, we rely on two limit order book metrics at the stock level obtained from the monthly Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services (CRSP) and aggregated across firms (equally-weighted). Columns I and II show that a higher information load is associated with a higher bid-ask spread (SPR) and effective spread (EFFSPR)—spread adjusted with the trading price—suggesting that excess information is related to information asymmetries.

Second, when the information load is high, estimated parameters of future returns or cash flows are more likely to be wrong, increasing estimation risk (see, for example, Coles and Loewenstein, 1988; Coles et al., 1995). To proxy the estimation risk, we consider analysts' forecast errors using the Institutional Brokers' Estimate System (Refinitiv, IBES North American Summary & Detail Estimates, Level 2, Current & History Data, Adjusted and Unadjusted)—I/B/E/S summary database. For a given firm and month, we calculate the averages of the absolute deviation of the mean, highest, and lowest EPS forecast from the actual value across analysts (FERROR1, FERROR2, and FERROR3, respectively). We then calculate the equally weighted cross-sectional averages. Columns III through V show that a high information load is associated with higher forecasting errors, likely exacerbating agents' confusion about firm performance. In columns VI and VII we include two metrics from the Michigan Surveys of Consumers: variations in business conditions' expectations and standard deviation of price level expectations for the next 12 months. We document a positive correlation between the heterogeneity among households on their expectations of business conditions and prices and the information load.

Finally, the information load is expected to increase in times of deteriorated financial or economic conditions. Columns VIII and IX show that we observe a boost in the quantity of news in periods of increased financial stress, measured by the Chicago Fed National Financial Conditions Index (NFCI) and St. Louis Fed Financial Stress Index (STLFSI4) from FRED economic data.

3 Motivating framework

As a motivating framework for our empirical analysis, we consider a discrete-time learning model. Our model incorporates a limited processing capacity of agents that is affected by information overload. In periods of excessive information load, it is more difficult for investors to process relevant information—an argument with empirical basis from the neuroscience literature (Reutskaja et al. 2018; Callicott et al. 1999; Jaeggi et al. 2007) and with theoretical support from psychologists and cognitive scientists (Miller 1956; Schroder et al. 1967; Simon and Newell 1971).

In our model, there are two periods ($t = 0, 1$) and a representative investor, who invests in a single risky asset at $t = 0$ by maximizing her CARA form utility function. The asset pays ν at $t = 1$, where the investor consumes the realized payoffs. The value of ν is normally distribute with $\nu \sim \mathcal{N}(\bar{\nu}, 1/\tau_0)$. There are Q outstanding shares of the asset. For simplicity, we assume that the risk-free rate is zero.

3.1 The learning process

At $t = 0$, the investor receives n news articles (i.e., n public signals) with information on the asset value. Let s_i be the information extracted from each news article $i \in [1, n]$:

$$s_i \quad = \quad \nu + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, 1/\tau_{news}), \tag{3}$$

in which, the noise components, ε_i, are independent and normally distributed, while τ_{news} represents the precision of s_i.

The investor has a limited processing capacity that is affected by the information load. Thus, in periods of excessive information flow, she cannot process all news.

Let m be the total number of news articles that the investor can process, with $m \leq n$. Using the standard Bayesian updating process, the following lemma characterizes the investor's posterior beliefs.

Lemma 1. *The posterior beliefs of the investor at $t = 0$ are normally distributed, and denoted by $\nu^{PostBlf} \sim \mathcal{N}(\hat{\nu}, 1/(\tau_0 + m\tau_{news}))$, where the expected value of their posterior beliefs is:*

$$\hat{\nu} \;=\; \frac{\tau_0}{\tau_0 + m\tau_{news}}\bar{\nu} + \frac{m\tau_{news}}{\tau_0 + m\tau_{news}}\bar{s}_m \;\; with \;\; \bar{s}_m = \frac{1}{m}\sum_{i=1}^{m} s_i. \tag{4}$$

Proof. See Appendix.

3.2 Information overload and the investor's processing capacity

Let M_{max} be the threshold number of news articles that can be processed without experiencing information overload. Then, information overload is defined as:

$$InfOver = \begin{cases} n - M_{max} & n > M_{max} \\ 0 & otherwise. \end{cases} \tag{5}$$

If the investor has an unlimited processing capacity (when $M_{max} \to \infty$), she would not be affected by information overload and process all news. Nevertheless, under a limited processing capacity, the investor can process all news only when the number of news articles is lower than M_{max}. Indeed, the current neuroscience and psychological literature document a hump-shaped relationship between information load and the investor' cognitive processing.[5] Following Gunaratne et al.

[5]By using functional magnetic resonance imaging, Reutskaja et al. (2018) examine the brain activity of the striatum and anterior cingulate cortex regions, which are in charge of attention allocation, reinforcement, decision-making, reward perception, and motivation. They show the

(2021) and Gunaratne et al. (2020), we characterize the number of news articles that the investor can actually process, m, by:

$$m = \begin{cases} M_{\mathit{InfOver}} & n > M_{max} \quad (or \quad equivalently, \quad InfOver > 0) \\ n & otherwise, \end{cases} \tag{6}$$

with

$$M_{\mathit{InfOver}} = \begin{cases} M_{max} - InfOver^{\alpha} & M_{max} > InfOver^{\alpha} \\ 0 & otherwise. \end{cases} \tag{7}$$

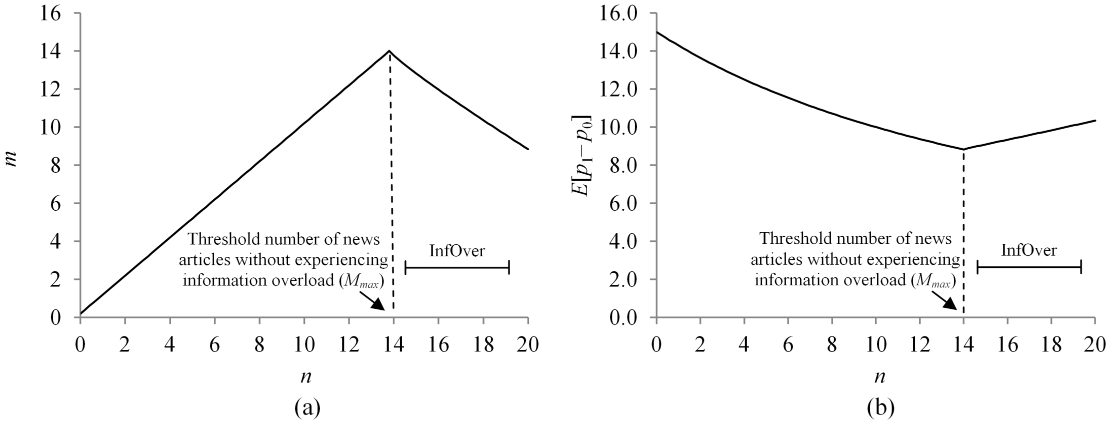
Figure 2, panel (a), visualizes the relationship between the number of news articles the investor receives (n) and the number of news articles that the investor can *actually* process, m. If $n \leq M_{max}$, there is no information overload, and thus the investor can process all n news. However, when $n > M_{max}$, the magnitude of such excess information induces an *information loss* in the investor's learning process ($\mathit{InfOver}^{\alpha}$), where α is the rate of loss with $0 \leq \alpha \leq 1$. Thus, in the InfOver region, m is negatively related to the amount of information overload through a power-law relationship of exponent α.

---

striatum and anterior cingulate cortex resembled an inverted V-shaped activity as a function of choice set size. This is because the brain unconsciously realizes that the cost outweighs the benefits of the cognitive processes after a threshold in the task cognitive load, where costs involve frustration, an overwhelming feeling, postponing important tasks of everyday life, and energy expending that can be used for the well functioning of the human body. Callicott et al. (1999) and Jaeggi et al. (2007) also show that the dorsolateral prefrontal cortex (which is in charge of cognitive flexibility and working memory) has an inverted U-shaped activity as a function of the cognitive load. Similarly, in the psychological literature, Schroder et al. (1967) present a model where the task performance of a decision-maker initially improves as more information is received. But, when the amount of information reaches a threshold, the additional information diminishes the quality of the decision-making process.

Figure 2: The relationship between information load, the investor's procession capacity, and the expected price change

Panels (a) shows the relationship between the number of news articles the investor receives, n, and the actual number of articles she can process, m. Panel (b) shows the relationship between m and the expected price change, $E[p_1 - p_0]$. Specifically, we plot the results presented in equations (6) and (10) by assuming $M_{max} = 14$, $\alpha = 0.9$, $\tau_{news} = 0.05$, $\tau_0 = 1$, $Q = 300$, $\theta = 0.1$, $\bar{\nu} = 50$.



3.3 Information overload and asset returns

At $t = 0$, after receiving (and learning from) the set of news articles, the investor maximizes her utility function based on the posterior beliefs about the asset value. Because the utility function takes the CARA form and all stochastic variables are normally distributed, the investor's optimization reduces to the usual mean-variance problem (see Ingersoll, 1987):

$$\max_y E_t[\eta] - \frac{\theta}{2} Var[\eta]. \tag{8}$$

Here, we assume a constant level of risk aversion θ, while η is the investor's net worth at $t = 0$, with $\eta = (\nu^{PostBlf} - p_0)y$, where $\nu^{PostBlf}$ reflects the investor's posterior beliefs about the asset value at $t = 1$, characterized in Lemma 1. p_0 is the asset price at $t = 0$ and y denotes the number of shares the investor holds.

Proposition 1 follows from solving the investor's optimization problem in (8) and using the market clearing condition of $y = Q$:

Proposition 1. *The equilibrium asset price at $t = 0$ is given by:*

$$p_0 = \frac{\tau_0}{\tau_0 + m\tau_{news}}\bar{\nu} + \frac{m\tau_{news}}{\tau_0 + m\tau_{news}}\bar{s}_m - \frac{\theta Q}{\tau_0 + m\tau_{news}} \tag{9}$$

and the expected asset price change is:

$$E[p_1 - p_0] = E[v] - E[p_0] = \frac{\theta Q}{\tau_0 + m\tau_{news}}. \tag{10}$$

Proof. See Appendix.

Note that, the risk premium for holding the asset, $\theta Q/(\tau_0 + m\tau_{news})$, is negatively related to the number of news articles that the investor can actually process, m, and the precision of the posterior beliefs.

Figure 2, panel (b), visualizes the relationship between the number of news articles the investor receives and the expected price change. When $n \leq M_{max}$, the investor does not suffer from information overload and she processes all information, increasing the precision of the posterior beliefs. Thus, the investor asks for a lower compensation to hold the risky asset (i.e., the risk premium is reduced).

Conversely, under information overload (i.e., when $n > M_{max}$), the information loss in the investor's learning process increases with information overload, reducing the precision of the investor's posterior beliefs, and making asset valuation more difficult, increasing estimation risk. Therefore, the investor asks for a larger compensation to hold the risky asset, which raises the value of $E[p_1 - p_0]$.

4 Empirical Analysis

4.1 Econometric model

To examine the predictive power of information load over future monthly market returns, we rely on the following regression model:

$$rx_{t+1}^{m} = \alpha_1 + \alpha_2 InfOver_t(tr) + \alpha_4 InfLoad_t + Controls + \varepsilon_{t+h}, \tag{11}$$

where rx_{t+1}^{m} is the one-period ahead market returns in excess of the risk-free rate. Market returns are based on the CRSP NYSE/AMEX/NASDAQ value-weighted portfolio in excess of the one-month Treasury bill rates and are obtained from Kenneth French's online data library.

We define $InfOver_t(tr)$ as the information load in month t above a threshold tr. We consider the one-standard-deviation band as the threshold, calculated using the twelve-month moving window sizes. One can think of this threshold as the maximum number of news that can be processed without experiencing information overload (corresponding to $Mmax$ in equation 5). Specifically:

$$InfOver_t(tr) = \begin{cases} InfLoad_t - tr_t & \text{if} \qquad InfLoad_t \geq tr_t \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The coefficient of interest is α_2. We hypothesize a non-linear relationship between information flow and the market risk premium as predicted by our model (see Figure 2). Because agents learn from news, information arrival would be beneficial, however only up to a certain threshold. When information flows exceed the threshold, investors' processing capacities will be exhausted, and they are less likely to process information correctly, increasing the market risk premium.

We control for the change in market sentiment (ΔSENT_t) as it has been shown to be a predictor of financial and economic activity (Tetlock, 2007; Garcia, 2013; van Binsbergen et al., 2022, among others). We define sentiment following Garcia (2013): For each day, we count the total number of positive and negative words as well as the total number of words in the corresponding lead paragraphs to obtain the proportion of positive and negative words. Market sentiment is then the difference between those proportions.

We also control for a set of variables that are shown to be significant predictors of market returns in traditional asset pricing studies. First, we consider the S&P 500 monthly dividend yield (DY_t) following Shiller (1978); Campbell (1987); Fama and French (1988), among others. We obtain data from Global Financial Data, Inc., GFDatabase (GFD). Second, we include realized volatility (RVOLA_t), calculated as the standard deviation of daily stock market returns, based on the CRSP NYSE/AMEX/NASDAQ value-weighted portfolio and obtained from Kenneth French's online data library.

Third, we include the Amihud's (2002) illiquidity measure, ILLIQ_t using price and volume data from CRSP. We also consider the default spread (DS_t), the term spread (TS_t), and the change in short-term interest rates ($\Delta\mathrm{STIR}_t$) following Keim and Stambaugh (1986); Campbell (1987); Fama and French (1989). DS_t is measured as the difference between Moody's Seasoned BAA and AAA corporate bond yields and data are retrieved from FRED, Federal Reserve Bank of St. Louis (DBAA and DAAA, respectively). TS_t is calculated as the difference between the 10-year Treasury bond and the 3-month T-bill yields, and $\Delta\mathrm{STIR}_t$ is the changes in the 3-month T-bill rate. We obtain interest rate data from the GFD. Finally, we include NBER recession dates to control for business cycles as news flows are expected to increase during periods of uncertainty.

In Table 2, we present summary statistics. On average, about 500 financial markets-related articles are printed at *the New York Times*, in a month, reaching a maximum of over 1000, making roughly one-third of the printed business articles. *InfOver* is correlated by about 47% with *InfLoad*, and 18% and 11% with DS_t and DY_t, respectively, suggesting that information overload is not likely to share common information with standard predictors of stock market returns. The Phillip-Perron stationarity tests strongly reject the null of the unit root for all of the variables.

4.2 Effects of information overload on stock returns: Time-series analysis

Table 3 presents the estimated coefficients from (11). In Column I, we include *InfLoad* as the main independent variable to examine the effects of information load on future market returns. We find that information load is not statistically related to stock market returns, perhaps not surprising given our prediction that the information load has a nonlinear relationship with stock market returns (see Figure 2, panel b).

To incorporate such nonlinearity, in column II, along with *InfLoad* and other standard predictors of returns, we include "excess" information load (one standard deviation above the historical mean)—*InfOver*. *InfOver* predicts higher next-period returns. The relationship is economically meaningful: a one standard deviation increase in information overload increases the market risk premium by almost 60 basis points. When controlled with *InfOver*, *InfLoad* reduces the market risk premium, suggesting that investors learn from news and their decision quality improves with information flow, but only up to a threshold.

In Column III, we further consider the "quality" of news in addition to the quantity and include consistency in the tone of the article, calculated as squared sentiment (SENT2). The higher the SENT2, the more consistent the articles on average are because the tone is becoming more and more optimistic or pessimistic. Apriori, we expect that when the news becomes inconsistent, investors are more likely to be confused and their information processing capacity might be exhausted well before the full quantity of information is used. We find that the quantity matters the most in predicting market returns.

In the baseline specifications, we cover a very long time series from July 1926 to December 2022, which is subject to significant changes both in terms of stock market and information flow dynamics. We then investigate the effects of information overload on stock market returns during different periods in columns IV and V. The economic impact of *InfOver* on stock market returns is more than doubled during the pre-WWII period compared to post-WWII. This result is intuitive, given that during the early period, printed newspapers were one of the main news outlets, and *The New York Times* extended its breadth and reach. Furthermore, the relationship is statistically significant both for the pre- and post-1995. Nonetheless, the explanatory power of *InfOver* on stock market returns gets weaker after the mid-1990s with the introduction of `NYTimes.com`, other news outlets, and the more common usage of the internet and social media.

Finally, we test the explanatory power of information overload on future market returns in comparison to other text-based measures: SENT, EPU, GPR, and NVIX. Columns VI and VII show that *InfOver* has about 1% of incremental explanatory power on market returns beyond the standard predictors and other text-based measures. None of the news-based measures have a statistically significant relationship with excess returns. These results are arguably expected given that both EPU and GPR are shown to be useful predictors of macroeconomic series rather than financial ones.

We also calculate the contribution of each regressor to the overall R^2 (share of explained variance) using Shapley value analysis to judge the relative importance of the variables in driving the changes in stock market returns. The standard predictors of returns (dividend yield, realized volatility, default spread, term spread, changes in interest rates, market liquidity, and NBER recession dates), together explain about 81% of the variation in the stock market returns. Yet, information load and overload explain about 13% of the variation, whereas the other news-based measures together explain just above 5% of the variation, with the contribution of EPU being the highest.

4.3 Effects of information overload on stock returns: Endogeneity concerns

So far, we have shown that information overload predicts stock market returns. However, such arguments assume that the supply of news is exogenous. Nonetheless, instead of information overload leading to an increase in stock market returns, unexpected changes in stock markets may lead to an increase in the number of news. Furthermore, omitted variables can affect both information load and market returns simultaneously. For example, worsening economic conditions increase the supply of information and also market risk premium.

We employ two credible attempts to address these endogeneity concerns. First, we use two-stage least squares (2SLS) estimation with the total number of articles related to obituaries, fashion, gaming, health, arts, travel, and home news as an instrument.[6] We choose these alternative news types because they are very unlikely to directly affect stock market returns, but their information load is likely to be correlated with financial market news, both driven by editorial choices or the structure of the newspaper at the time. Furthermore, reverse causality is limited in this setting, because changes in stock market dynamics are unlikely to affect obituaries news, for instance.

[6] As a robustness, we repeat the analysis using different subsets of the news, such as including news only from obituaries, fashion, gaming, and home. We reach similar conclusions using different newsgroups.

To generate the instrumental variables (IVs), we follow our main methodology for calculating information load and overload using this alternative set of news instead of financial market news, denoted by $\mathit{InfLoad}^{IV}$ and $\mathit{InfOver}^{IV}$. In columns I and II of Table 4, we report the first-stage regressions for both IVs. There is a positive and significant relationship between $\mathit{InfOver}^{IV}$ and *InfOver* and $\mathit{InfLoad}^{IV}$ and *InfLoad*. The $F-$statistics are 23 and 67, respectively, both significantly greater than 10, supporting the relevancy of the IVs. The second stage regression (column III) provides supporting evidence on the causal effect of information overload on the next period stock market returns.

As a second approach to address the endogeneity concerns, we examine the effects of information overload on stock market returns by considering plausible exogenous shifters of information load causing a *reduction* in the amount of news published in *the New York Times*. For example, we include newspaper strikes, changes in the physical size of the print edition, the editorial decisions regarding the combination of some sections in the newspaper, or when `NYTimes.com` began publishing and run the following regression:

$$\begin{aligned} rx^m_{t+1} &= \alpha_1 + \alpha_2 \mathit{InfOver}_t(tr) + \alpha_3 \mathit{InfOver}_t(tr) \times D_{i,t} + \alpha_4 D_{i,t} \\ &+ \alpha_5 \mathit{InfLoad}_t + Controls + \varepsilon_{t+1}, \end{aligned} \tag{13}$$

where $D_{i,t}$ is a dummy variable that equals 1 up to six months after the start of each event i and 0 otherwise. When an exogenous event reduces the supply of news, and hence reduces information overload, we expect the effects of *InfOver* on market returns to be lower. Columns IV and V of Table 4 report the results. We find that higher information overload continues to predict higher stock market returns ($\widehat{\alpha_2} + \widehat{\alpha_3} > 0$) but with significantly reduced sensitivity ($\widehat{\alpha_3} < 0$).

Overall, both analyses confirm our conjecture and increase our confidence in the main finding that investors require a higher risk premium to hold the market portfolio after periods of excessive information load.

4.4 Effects of information overload on stock returns: Cross-sectional analysis

We expect *InfOver* to have cross-sectional effects on stock returns based on firm characteristics. It is because, in periods of information overload, investors are cognitively constraint and it is optimal for them to allocate their processing capacities into a certain group of stocks that requires less attention, such as larger stocks or the constituents of a stock market index (see, the category learning model of Peng and Xiong, 2006). Barberis et al. (2005) show that, after a firm is excluded from the Standard and Poor's (S&P500) index, it attracts less investor attention. Moreover, stocks with weaker arbitrage forces (such as smaller or riskier stocks) are prone to speculative mispricing as they tend to be more difficult to value (e.g., Baker and Wurgler, 2006). Finally, Akbas et al. (2018) show that information processing is more difficult for stocks excluded in the S&P500 index and stocks with lower institutional ownership as retail investors have fewer resources compared to institutional investors. Thus, we expect information overload to exacerbate the mispricing of these stocks by making valuation even more difficult and increasing estimation risk.

Overall, we conjecture that following periods of information overload, investors would require a higher risk premium to hold smaller and more volatile stocks, stocks that are not included in S&P500 index, and with lower institutional ownership. To examine these conjectures, we run the following regression model:

$$\begin{aligned} r_{i,t+1} &= \alpha_1 + \alpha_2 InfOver_t(tr) \times D_{i,t}^{Grp1} + \alpha_3 InfOver_t(tr) \times D_{i,t}^{Grp2} \\ &+ \alpha_4 InfLoad_t + Controls + \varepsilon_{t+h}, \end{aligned} \tag{14}$$

where $r_{i,t+1}$ is the stock return of firm i in month $t + 1$. $D_{i,t}^{Grp1}$ is a dummy variable that equals to 1, if in month t, stock i is small and highly volatile, is not a constituent of the S&P500 index, has a smaller share of institutional ownership in shares outstanding. Analogously, $D_{i,t}^{Grp2}$ is a dummy variable that equals to 1, if in month t, stock i is large and less volatile, is a constituent of the S&P500 index, and has a higher share of institutional ownership. We measure size and volatility using the market capitalization and monthly standard deviation of daily returns, respectively. High and low values are based on the 20% top and 20% bottom percentiles of the corresponding firm characteristics. We obtain data from CRSP. We get the share of institutional ownership through Thomson Reuters Stock Ownership data for more than 80,000 firms spanning from March 1983 to December 2022. The data are based on the SEC's Form 13F, which are filed quarterly by large institutional investment managers.

Table 5 reports the estimated coefficients from (14). Column I shows that, unconditionally, information overload not only leads to higher stock market returns (as documented so far), but also higher individual stock market returns. Columns II-VI show that smaller and more volatile stocks, and stocks that are not included in the S&P500 index, the ones with lower institutional ownership exhibit significantly higher returns compared to their counterparts during periods of information overload. Thus, the results confirm our conjecture that the effects of information overload on next-period returns depend on firm characteristics and that information overload exacerbates mispricing by aggravating investors' capacity constraints.

4.5 Effects of information overload on stock returns: Clustering earning announcements

This section aims to confirm our main findings in a different setting while relying on a different proxy of information overload. Specifically, we use clustered earnings announcements as another measure of information overload for investors and analysts and study their effects on firm-specific returns.

Earnings announcements provide new information to market participants regarding the health and performance of a firm. Not surprisingly, there is a vast literature that studies the effects of firms' earnings announcements on financial markets (see Aharony and Swary, 1980; Easton and Zmijewski, 1989; Savor and Wilson, 2016, among others). Also, earnings announcements cluster. For instance, the Securities and Exchange Commission (SEC) requires U.S. publicly traded firms to disclose their earnings within a few weeks after the end of each fiscal quarter. In our sample, almost two-thirds of the firms disclose their year-end earnings during February.

We claim that, when many firms release earnings reports simultaneously in a day—information overload— it becomes challenging for investors to focus on each announcement adequately. Consequently, the effects of earnings announcements on returns should depend on the number of firms making these announcements.

To test our hypothesis, we collect earnings announcement data from I/B/E/S made by 15,463 firms, spanning from January 1985 to December 2022. We focus only on positive announcements when the actual earnings exceed the average EPS forecasts for a firm in a given forecast period. We then split data into quintiles depending on the number of firms announcing their earnings on the same day.[7] Specifically, Group 1 corresponds to the lowest quintile, i.e., considers the days with the smallest number of firms making announcements, while Group 5 corresponds to the highest quintile. We then create five dummy variables, D_t^j, which is equal to 1 if the number of earning announcements on day t falls into Group j, and run

[7] To account for the changing number of "active" firms throughout the years, we also split data based on the *frequency* of firms announcing their earnings on the same day, reaching similar results.

the following regression:

$$r_{i,t} = \sum_{j=1}^{5} \alpha_j D_t^j + \beta r x_t^m + \gamma esurp_{i,t}|_{esurp_{i,t}>0} + \varepsilon_{i,t}, \tag{15}$$

where $r_{i,t}$ is the return of firm i making an announcement on day t, rx_t^m is the market excess return, and $esurp_{i,t}$ is the earning surprise.

Figure 3 plots the estimated coefficients of the groups ($\widehat{\alpha}$s) along with their 95% confidence intervals. All $\widehat{\alpha}$s are positive, suggesting that irrespective of the earning cluster group, the average returns are positive on the days of positive earnings announcements. Importantly, we find a monotonically decreasing relationship: the mean value of returns is higher for Group 1 (the bucket representing the days with the smallest number of firms announcing) in comparison to Group 5 (the bucket representing the days with the largest number of firms announcing), all else equal. Thus, on the days in which earning announcements cluster—investors face information overload—the average returns are significantly less in comparison to the days in which fewer firms are announcing.

4.6 Robustness

To test the sensitivity of our findings, we run several robustness tests. We start by altering the definition of information overload by either using different thresholds or different estimation windows. First, instead of calculating information overload using the deviation over a one-standard-deviation band, we measure the information overload via the deviations from alternative thresholds: the historical mean and trend estimated via the Hamilton (2017) filter. Second, instead of using 12-month rolling window to calculate the one-standard deviation threshold, we employ moving window sizes of three, six, and 24 months.

Third, instead of calculating stock market volatility as the standard deviation of

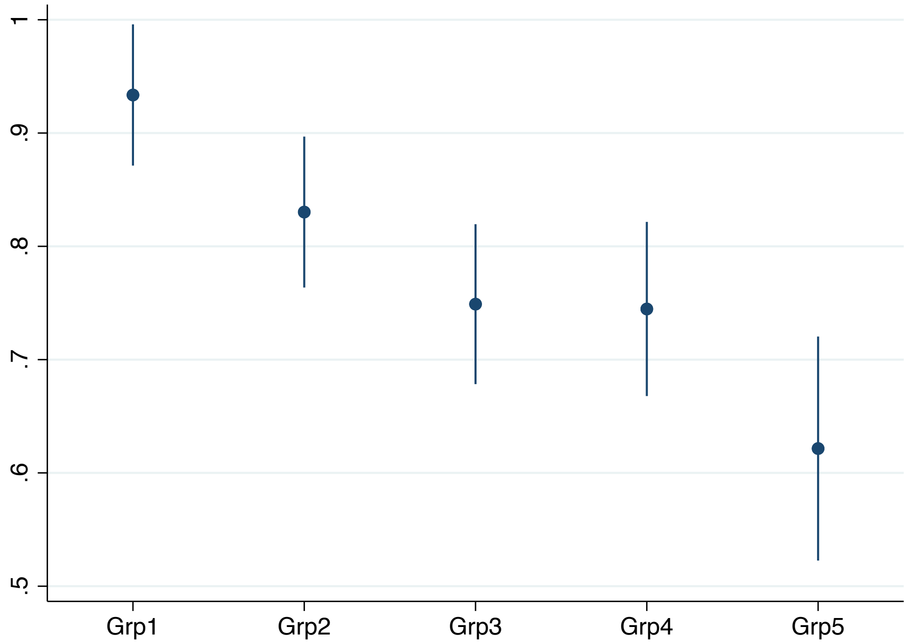


Figure 3: The effects of clustered earnings on stock returns

The figure presents the estimated coefficients of groups ($\widehat{\alpha}$s) in (15) along with their 95% confidence intervals. Group 1 corresponds to the lowest quintile, i.e., considers the days with the smallest number of firms making announcements, while Group 5 corresponds to the highest quintile. Earnings announcements are from I/B/E/S from January 1985 to December 2022. Stock return data are obtained from the daily Center for Research in Security Prices, CRSP 1925 US Indices Database, and Wharton Research Data Services.

daily stock returns in a given month, we estimate it via a GARCH(1,1) model (11).

Fourth, we set the maximum lag order of autocorrelation to 6 instead of 2 while calculating Newwey-west standard errors, following the rule of thumb in Greene (2000).

Fifth, in addition to the baseline control variables, we include the inflation rate, calculated as the log changes in the U.S. consumer price index, changes in the industrial production index, and market risk perception proxied by the Duration of Low Risk (DLR) of Danielsson et al. (2023).

Finally, we test whether our results are driven by extreme events. To this end, we exclude stock market crashes, identified as the periods in which the S&P 500 index decrease over 10%. In addition, we exclude major global episodes: the Great Depression (1929–1935), the Global Financial Crisis (2007–2009), and World Wars (1914–1918, 1939–1945).

Table 6 presents the results. We conclude that our findings are robust to the specifications we consider and that information overload significantly explains the next period excess market returns.

5 Conclusion

In this paper, we study the role of media news in stock markets and provide the first evidence that the excessive amount of information predicts stock returns. We structure our empirical analysis around a discrete-time learning model, which links information load with asset prices when the investor is attention-constrained. Using textual analysis and machine learning tools, we compute a news-based historical information load index by considering over two million articles printed in *The New York Times* from January 1, 1885, to December 31, 2022. We use these novel data to study the effects of information load on stock market dynamics.

We quantify information overload as the deviation of the number of financial markets articles printed in a given month from a historical threshold. We argue that this threshold can be thought of as the representative agent's processing capacity limits. We find that information flow is valuable and agents learn from news. However, the relationship between information flow stock returns is non-linear. Following excessive news-flow periods, investors require a higher risk premium to hold the risky asset. We also document the cross-sectional effects of information overload on individual firm-level stock returns based on different firm characteristics. Such findings are consistent with cognitive theories emphasizing that information overload exhausts investors' processing capacities, deteriorates their decision accuracy, and thus, increases information and estimation risk.

Appendix

Proof of Lemma 1. The proof follows the standard Bayesian updating process of signals that are independent and normally distributed with known precisions.

Proof of Proposition 1. The proof follows from obtaining the first-order condition of (8), and using the market clearing condition. The proof for the expected price change follows from substituting the asset prices at $t = 0$ and $t = 1$ inside the expectation operator in equation (10); and then using equations (3) and (9).

References

Aharony, J. and I. Swary (1980). Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis. *The Journal of finance 35*(1), 1–12.

Akbas, F., S. Markov, M. Subasi, and E. Weisbrod (2018). Determinants and consequences of information processing delay: Evidence from the thomson reuters institutional brokers' estimate system. *Journal of Financial Economics 127*(2), 366–388.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*, 31–56.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 61*, 1645–1680.

Baker, S., N. Bloom, and S. Davis (2016). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Quarterly Journal of Economics 131*, 1593–1636.

Barberis, N., A. Shleifer, and J. Wurgler (2005). Comovement. *Journal of financial economics 75*(2), 283–317.

Bawden, D. and L. Robinson (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of information science 35*(2), 180–191.

Blair, A. (2012). Information overload's 2,300-year-old history. Harvard business review online resources.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment 2008*(10), P10008.

Caldara, D. and M. Iacoviello (2022). Measuring geopolitical risk. *American Economic Review 112*(4), 1194–1225.

Callicott, J. H., V. S. Mattay, A. Bertolino, K. Finn, R. Coppola, J. A. Frank, T. E. Goldberg, and D. R. Weinberger (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional mri. *Cerebral Cortex 9*(1), 20–26.

Calomiris, C. W. and H. Mamaysky (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics 133*(2), 299–336.

Campbell, J. (1987). Stock returns and the term structure. *Journal of Financial Economics 18*, 373–399.

Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics 70*(2), 223–260.

Cohen, L. and D. Lou (2012). Complicated firms. *Journal of financial economics 104*(2), 383–400.

Coles, J. and U. Loewenstein (1988). Equilibrium pricing and portfolio composition in the presence of uncertain parameters. *Journal of Financial Economics 22*, 279–303.

Coles, J., U. Loewenstein, and J. Suay (1995). On equilibrium pricing under parameter uncertainty. *The Journal of Financial and Quantitative Analysis 30*, 347–374.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

Danielsson, J., M. Valenzuela, and I. Zer (2023). The impact of risk cycles on business cycles: a historical view. *The Review of Financial Studies 36*(7), 2922–2961.

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

Easton, P. D. and M. E. Zmijewski (1989). Cross-sectional variation in the stock market response to accounting earnings announcements. *Journal of Accounting and economics 11*(2-3), 117–141.

Edmunds, A. and A. Morris (2000). The problem of information overload in business organisations: a review of the literature. *International journal of information management 20*(1), 17–28.

Eppler, M. and J. Mengis (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines. *The Information Society 20*, 325–344.

Fama, E. and K. French (1988). Dividend tields and expected stock returns. *Journal of Financial Economics 22*, 3–25.

Fama, E. and K. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*, 23–49.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance 68*, 1267–1300.

Gleick, J. (2011). *The information: A history, a theory, a flood.* Vintage.

Greene, W. (2000). *Econometric analysis.* Prentice Hall.

Grisé, M.-L. and R. B. Gallupe (1999). Information overload: Addressing the productivity paradox in face-to-face electronic meetings. *Journal of Management Information Systems 16*(3), 157–185.

Gunaratne, C., N. Baral, W. Rand, I. Garibay, C. Jayalath, and C. Senevirathna (2020). The effects of information overload on online conversation dynamics. *Computational and Mathematical Organization Theory 26*, 255–276.

Gunaratne, C., W. Rand, and I. Garibay (2021). Inferring mechanisms of response prioritization on social media under information overload. *Nature Scientific Reports 11*(1), 1346.

Hamilton, J. D. (2017). Why you should never use the hodrick-prescott filter. *The Review of Economics and Statistics forthcoming.*

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hong, H. and D. A. Sraer (2016). Speculative betas. *The Journal of Finance 71*(5), 2095–2144.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance 54*(6), 2143–2184.

Hong, H. and J. C. Stein (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies 16*(2), 487–525.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making.* Maryland: Rowman & Littlefield.

Jaeggi, S. M., M. Buschkuehl, A. Etienne, C. Ozdoba, W. J. Perrig, and A. C. Nirkko (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience 7*(2), 75–89.

Johnston, J. C. and H. Pashler (1998). Attentional limitations in dual-task performance. *Attention*, 155–189.

Kahneman, D. (1973). *Attention and effort*, Volume 1063. Citeseer.

Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics 17*, 357–390.

Lee, Y.-J. (2012). The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research 29*(4), 1137–1170.

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *Journal of Finance 69*, 1643–1671.

Manela, A. and A. Moreira (2017). News implied volatility and disaster concerns. *Journal of Financial Economics 123*(1), 137–162.

Miller, J. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review 63*, 81–97.

Muslu, V., S. Radhakrishnan, K. Subramanyam, and D. Lim (2015). Disclosures and the information environment. *Management Science 61*, 931–948.

Peng, L. and W. Xiong (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics 80*(3), 563–602.

Reutskaja, E., A. Lindner, R. Nagel, R. A. Andersen, and C. F. Camerer (2018). Choice overload reduces neural signatures of choice set value in dorsal striatum and anterior cingulate cortex. *Nature Human Behaviour 2*(12), 925–935.

Roetzel, P. G. (2019). Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research 12*, 479–522.

Savor, P. and M. Wilson (2016). Earnings announcements and systematic risk. *The Journal of Finance 71*(1), 83–138.

Schroder, H. M., M. J. Driver, and S. Streufert (1967). *Human information processing—Individuals and groups functioning in complex social situations.* New York: Holt, Rinehart, & Winston.

Shiller, R. (1978). Stock prices and social dynamics. *Brookings Papers on Economic Activity 2*, 457–510.

Simon, H. and A. Newell (1971). Human problem solving: The state of the theory in 1970. *American Psychologist 26*, 145–159.

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*, 1139–1168.

van Binsbergen, J. H., S. Bryzgalova, M. Mukhopadhyay, and V. Sharma (2022). (almost) 200 years of news-based economic sentiment. *Available at SSRN 4261249*.

Table 1: Pearson correlation coefficients

In this table, we present contemporaneous Pearson correlation coefficients (ρ) of information load with proxies of information risk, estimation risk, and financial stress indexes. *InfLoad* is introduced in (2). The bid-ask spread (SPR) and effective bid-ask spread (EFFSPR) are stock-level metrics from the monthly WRDS CRSP database and are aggregated across firms (equally weighted). FERROR1, FERROR2, and FERROR3 are the average absolute deviations of the mean, highest, and lowest EPS forecast from the actual value. Data are from the I/B/E/S summary database. First, for a given firm and month, we calculate the average of absolute deviations of EPS forecast errors across analysts. We then calculate the equally weighted cross-sectional averages. devPX and devBEX are the deviations of price expectations and business conditions expectations, respectively, obtained from Michigan Surveys of Consumers. Finally, NFCI is the Chicago Fed National Financial Conditions Index and STLFSI4 is St. Louis Fed Financial Stress Index, obtained from FRED economic data.

	I	II	III	IV	V	VI	VII	VIII	IX
	SPR	EFFSPR	FERROR1	FERROR2	FERROR3	devPX	devBEX	NFCI	STLFSI4
ρ	0.49	0.45	0.28	0.31	0.33	0.46	0.12	0.24	0.21
p-values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	1161	918	468	468	468	536	536	609	336

Table 2: Summary statistics

This table reports the mean, minimum, maximum, standard deviation, and the $p-$values corresponding to the Philips-Perron stationarity test results of the main variables included in our analysis. The last row reports the series' correlation at the column header with information overload $InfOver_t$, which is introduced in (12). Monthly excess market returns, rx_t^m are the CRSP NYSE/AMEX/NASDAQ value-weighted portfolio returns in excess of the one-month treasury bill rate. $InfLoad_t$ is the number of financial markets-related news in a given month. ΔSENT_t is the first difference of the market sentiment measure, calculated as the difference between the proportion of positive and negative words, following Garcia (2013). DY_t is the S&P 500 monthly dividend yield, ΔSTIR_t is the change in three-month T-bill rates, TS_t is the term spread, calculated as the difference between the ten-year Treasury bond and the three-month Treasury bill yields. Default spread, DS_t, is measured as the difference between BAA and AAA corporate bond spreads, RVOLA_t is monthly realized volatility and calculated as the standard deviation of stock market returns. Finally, ILLIQ_t is the Amihud's illiquidity measure. All of the variables are monthly estimates from July 1926 to December 2022, where available. Data sources: ProQuest, TDM Studio. New York Times Historical Newspapers, Global Financial Data, Inc., GFDatabase, FRED, the Federal Reserve Bank of St Louis, Kenneth French's online data library, and the Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services.

	I	II	III	IV	V	VI	VII	VIII	IX	X
	rx_t^m	$InfOver_t$	$InfLoad_t$	ΔSENT_t	DY_t	ΔSTIR_t	TS_t	DS_t	RVOLA_t	ILLIQ_t
mean	0.68	14.10	481.47	-0.00	3.64	0.00	1.63	1.12	0.04	0.96
min	-29.13	0.00	41.08	-0.35	1.08	-3.85	-2.64	0.32	0.01	-81.17
max	38.85	470.54	1154.33	0.36	9.63	2.60	4.41	5.64	0.27	371.17
stdev	5.34	53.83	269.05	0.07	1.61	0.41	1.26	0.68	0.03	16.64
PP–stationarity	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01
corr(infover)	-0.02	1.00	0.47	0.04	0.11	0.03	-0.05	0.18	0.08	0.07

Table 3: Information overload and market excess returns
In this table, we report the estimated coefficients of the time-series regressions introduced in (11). The dependent variable is the one-period ahead market returns in excess of the risk-free rate. *InfLoad* and *InfOver* are introduced in (2) and (12), respectively. SENT2 is the squared sentiment measure introduced in Section 4.1. EPU is the Economic Policy Uncertainty index of Baker et al. (2016), GPR is the Geopolitical Risk Index of Caldara and Iacoviello (2022), and NVIX is the News-implied Volatility Index of Manela and Moreira (2017). The rest of the variables are described in Table 2. All of the variables are in monthly frequency from July 1926 to December 2022, where available, and standardized to ease the interpretation of the coefficients. Newey-West standard errors with the maximum lag order of autocorrelation of 2 are reported. ***, **, and * denote significance at the 1%, 5%, and 10% level (two-sided), respectively. Data sources: ProQuest, TDM Studio. New York Times Historical Newspapers, Global Financial Data, Inc., GFDatabase, the Federal Reserve Bank of St Louis, Kenneth French's online data library, and the Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services.

Dep. var.: rx^m_{t+1}	I	II	III	IV	V	VI	VII
$InfOver_t$		0.57***	0.58***				0.57**
		(0.219)	(0.221)				(0.239)
$InfLoad_t$	-0.17	-0.55**	-0.56*	-0.59**	-0.54**		-0.48
	(0.250)	(0.272)	(0.297)	(0.278)	(0.271)		(0.296)
SENT2_t			-0.02				
			(0.175)				
$InfOver_t \times D_t^{preWWII}$				0.58**			
				(0.263)			
$InfOver_t \times D_t^{postWWII}$				0.24**			
				(0.121)			
$InfOver_t \times D_t^{pre1995}$					0.56**		
					(0.220)		
$InfOver_t \times D_t^{post1995}$					0.19*		
					(0.108)		
ΔSENT_t	-0.30	-0.30	-0.31	-0.29	-0.31*	-0.21	-0.20
	(0.191)	(0.189)	(0.195)	(0.190)	(0.189)	(0.222)	(0.222)
EPU_t						0.25	0.18
						(0.247)	(0.249)
GPR_t						0.00	0.00
						(0.191)	(0.209)
NVIX_t						-0.03	0.01
						(0.212)	(0.203)
DY_t	0.48**	0.66***	0.66***	0.68***	0.68***	0.47**	0.64**
	(0.227)	(0.235)	(0.235)	(0.237)	(0.238)	(0.211)	(0.274)
ΔSTIR_t	-0.38*	-0.41**	-0.41**	-0.42**	-0.41*	-0.37*	-0.39*
	(0.214)	(0.207)	(0.207)	(0.205)	(0.207)	(0.218)	(0.212)
TS_t	0.14	0.18	0.18	0.19	0.19	0.09	0.15
	(0.171)	(0.170)	(0.170)	(0.169)	(0.169)	(0.181)	(0.180)
DS_t	0.51	0.44	0.45	0.41	0.45	0.45	0.42
	(0.426)	(0.427)	(0.434)	(0.427)	(0.426)	(0.435)	(0.454)
RVOLA_t	0.01	0.02	0.02	0.02	0.00	-0.18	-0.16
	(0.326)	(0.315)	(0.316)	(0.314)	(0.313)	(0.387)	(0.377)
ILLIQ_t	1.11***	1.08***	1.08***	1.08***	1.08***	1.10***	1.07***
	(0.307)	(0.296)	(0.296)	(0.292)	(0.296)	(0.312)	(0.301)
NBER	-1.62***	-1.81***	-1.80***	-1.80***	-1.80***	-1.44***	-1.67***
	(0.553)	(0.569)	(0.569)	(0.569)	(0.568)	(0.554)	(0.580)
Adj. R^2 (%)	5.37	6.45	6.37	6.52	6.46	5.46	6.42
No Obs.	1,150	1,150	1,150	1,150	1,150	1,055	1,055

Table 4: Information overload and market excess returns–Endogeneity concerns *InfLoad* and *InfOver* are introduced in (2) and (12), respectively. $InfLoad^{IV}$ and $InfOver^{IV}$ are instruments for *InfLoad* and *InfOver*, calculated using articles related to obituaries, fashion, gaming, health, arts, travel, and home instead of financial market news. Columns I and II report the first stage of the two-stage least squares (2SLS) regressions for both IVs. Column III reports the second-stage regression results. In columns IV and V, we report the estimated coefficients from (13), in which we consider exogenous events that reduce the supply of news. $D_{i,t}$ is a dummy variable that equals 1 up to six months after the start of each event i and 0 otherwise. The rest of the variables are described in Table 2. All of the control variables are included but not reported for the sake of brevity. All of the variables are in monthly frequency from July 1926 to December 2022, where available, and standardized to ease the interpretation of the coefficients. ***, **, and * denote significance at the 1%, 5%, and 10% level (two-sided), respectively. Data sources: ProQuest, TDM Studio. New York Times Historical Newspapers, Global Financial Data, Inc., GFDatabase, the Federal Reserve Bank of St Louis, Kenneth French's online data library, and the Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services.

	I	II	III	IV	V
	1st stage	1st stage	2nd stage	exo. events	exo. events
	InfOver	*InfLoad*			
$InfOver_t$			1.14**	0.57***	0.57***
			(0.474)	(0.220)	(0.219)
$InfOver^{IV}$	0.34***	0.04*			
	(0.029)	(0.021)			
$InfLoad_t$			-1.37	-0.56**	-0.56**
			(0.924)	(0.273)	(0.272)
$InfLoad^{IV}$	0.09**	0.22***			
	(0.037)	(0.027)			
$InfOver_t \times D_{i,t}$				-0.10**	-0.11***
				(0.043)	(0.025)
$D_{i,t}$				-0.35	
				(1.168)	
$F(10, 1139)$	23.52	66.40			
Controls	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	16.39	36.27	5.70	6.36	6.44
No Obs.	1.150	1.150	1,150	1,150	1,150

Table 5: Information overload and cross-section of returns

In this table, we report the estimated coefficients from (14). The dependent variable is $r_{i,t+1}$, stock returns of firm i in month $t+1$. *InfLoad* and *InfOver* are introduced in (2) and (12), respectively. $D_{i,t}^{gr1}$ is a dummy variable that equals to 1, if in month t, stock i is not a constituent of the S&P500 index, is small, is highly volatile, and has a smaller share of institutional ownership in shares outstanding. Analogously, $D_{i,t}^{gr2}$ is a dummy variable that equals to 1, if in month t, stock i is a constituent of the S&P500 index, is large, is less volatile, and has a higher share of institutional ownership. We measure size and volatility using the market capitalization and monthly standard deviation of daily returns, respectively. High and low values are based on the 20% top and 20% bottom percentiles of the corresponding firm characteristics. The rest of the variables are described in Table 2. All of the variables are in monthly frequency from July 1926 to December 2022, where available, and standardized to ease the interpretation of the coefficients. ***, **, and * denote significance at the 1%, 5%, and 10% level (two-sided), respectively. Data sources: ProQuest, TDM Studio. New York Times Historical Newspapers, Global Financial Data, Inc., GFDatabase, the Federal Reserve Bank of St Louis, Kenneth French's online data library, and the Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services, and Thomson Reuters Stock Ownership database.

Dep. var.: $r_{i,t+1}$	I baseline	II size	III volatility	IV S&P500	V inst. ownership
$\mathit{InfOver}_t$	0.46**				
	(0.185)				
$\mathit{InfOver}_t \times D_{i,t}^{gr1}$		0.34***	0.23**	0.45**	0.16**
		(0.098)	(0.109)	(0.183)	(0.080)
$\mathit{InfOver}_t \times D_{i,t}^{gr2}$		0.11*	0.11**	0.13***	0.07
		(0.060)	(0.046)	(0.050)	(0.078)
$\mathit{InfLoad}_t$	-0.59***	-0.43**	-0.41*	-0.59***	-0.15
	(0.226)	(0.216)	(0.215)	(0.226)	(0.309)
SENT_t	-0.10	-0.09	-0.09	-0.10	-0.18
	(0.188)	(0.188)	(0.188)	(0.188)	(0.220)
DY_t	0.73***	0.69***	0.68**	0.73***	0.41
	(0.266)	(0.265)	(0.265)	(0.266)	(0.631)
$\Delta\mathrm{STIR}_t$	-0.56***	-0.55***	-0.55***	-0.56***	-0.41*
	(0.188)	(0.189)	(0.189)	(0.188)	(0.234)
TS_t	0.18	0.15	0.15	0.18	0.13
	(0.216)	(0.215)	(0.215)	(0.216)	(0.246)
DS_t	0.91***	0.92***	0.92***	0.91***	0.84
	(0.314)	(0.315)	(0.315)	(0.314)	(0.614)
RVOLA_t	0.01	0.02	0.01	0.01	0.11
	(0.408)	(0.411)	(0.411)	(0.408)	(0.511)
ILLIQ_t	0.72***	0.72***	0.73***	0.72***	1.69
	(0.260)	(0.261)	(0.257)	(0.260)	(1.143)
NBER	-2.38***	-2.28***	-2.28***	-2.38***	-2.11
	(0.841)	(0.838)	(0.839)	(0.841)	(1.359)
Adj. R^2 (%)	0.96	0.93	0.92	0.96	0.97
No Obs.	4,666,461	4,662,778	4,665,535	4,666,461	2,648,547

Table 6: Robustness

In this table, we present the results of the robustness analysis. Column I reports the baseline specification. In columns II and IIII, we define the threshold as historical mean and trend estimated via the Hamilton (2017) filter. In columns IV to VI, we set 3, 6, and 24 months of rolling windows to calculate threshold, respectively, instead of using 12 months as in the baseline specification. In column VII, we estimate volatility through a GARCH(1,1) model. In column VIII, we alter the maximum lag length for the Newey-West standard error calculations. In column IX, we include other macro controls (changes in industrial production index, CPI inflation, and DLR of Danielsson et al. (2023) as a proxy of risk appetite). In column X, we exclude major stock market crashes, defined as when the S&P500 index drops over 10%. Finally, in column XI, we exclude major global episodes: the Great Depression (1929–1935), the Global Financial Crisis (2007–2009), and World Wars (1914–1918, 1939–1945). $InfOver_t$ is the information overload measure introduced in (12). The rest of the variables are described in Table 2. All of the explanatory variables are included but not reported for the sake of brevity. All of the variables are in monthly frequency from July 1926 to December 2022, where available, and standardized to ease the interpretation of the coefficients. Newey-West standard errors with the maximum lag order of autocorrelation of 2 are reported, except in column VIII. ***, **, and * denote significance at the 1%, 5%, and 10% level (two-sided), respectively. Data sources: ProQuest, TDM Studio. New York Times Historical Newspapers, Global Financial Data, Inc., GFDatabase, the Federal Reserve Bank of St Louis, Kenneth French's online data library, and the Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Specification:	baseline	mean	hamilton	3months	6months	24 months	Garch	Newey-lags	Macro	No crashes	No crises
$InfOver_t$	0.57***	0.78***	0.66***	0.50**	0.50**	0.52***	0.56**	0.57***	0.57**	0.49***	0.38**
	(0.219)	(0.251)	(0.201)	(0.209)	(0.209)	(0.174)	(0.240)	(0.213)	(0.227)	(0.181)	(0.155)
$InfLoad_t$	-0.55**	-0.91**	-0.79**	-0.47*	-0.48*	-0.55**	-0.55**	-0.55**	-0.61**	-0.45*	-0.26
	(0.272)	(0.354)	(0.337)	(0.274)	(0.269)	(0.280)	(0.275)	(0.249)	(0.301)	(0.265)	(0.207)
									(0.156)		
controls											
Adj. R^2 (%)	6.45	6.78	6.34	6.24	6.23	6.15	6.46	6.45	6.83	5.21	2.74
No Obs.	1,150	1,150	1,150	1,150	1,150	1,150	1,140	1,150	1082	1118	946