# Learning from news, information flow, and financial markets *

Alejandro Bernales Marcela Valenzuela Ilknur Zer

October 2024

**Abstract**

Neuroscience reveals that human brain activity and cognitive load exhibit an inverted V-shape relationship, showing a turning point, beyond which agents are unable to process additional information. We investigate the implications of this cognitive phenomenon within the financial markets, where participants are continually faced with and have to process news. We use textual analysis and machine learning tools applied to *The New York Times* since 1885 to measure information overload and find that investors learn from the news but only up to a threshold. Information overload exhausts investors' processing capacity, making the extracted information imprecise. Increased estimation risk, in turn, leads to higher market risk premium. Information overload also has cross-sectional effects: stocks that are difficult to value and require large learning efforts have higher expected returns.

**Keywords:** Information overload, media content, sentiment, limited attention, predicting returns, behavioral biases.

**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. This paper was previously circulated as "Effects of Information Overload on Financial Markets: How Much Is Too Much?" We thank Theis Ingerslev Jensen, Charles Calomiris, Paul Tetlock, Thierry Foucault, Dong Lou, Ioanid Rosu, Frank Warnock, Svetlana Bryzgalova, Ana Elisa Pereira, Consuelo Silva, Felipe Aldunate, Carlos Canon, George Skiadopoulos, Pablo Neudorfer, Richard Holowczak, Louisa Chen, and seminar participants at the Econdat 2024, 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, FWFS-GNY Symposium at Pace University, University of Chile, FEBS 2024, King's College London, Essex Business School, Santiago Finance Workshop 2024. 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 Federal Reserve System.

# 1 Introduction

*In an information-rich world, a wealth of information creates a poverty of attention —Nobel laureate Herbert Simon, 1971.*

Processing new information requires substantial resources, such as cognitive capability, time, or budget. In our daily lives, however, we have limited resources and are plagued with excess information. Indeed, early cognitive scientists have demonstrated that an "excessive" amount of information leads to cognitive overload, making it difficult for individuals to follow and process information accurately (Schroder et al., 1967; Miller, 1956; Simon and Newell, 1971). Along similar lines, in a neuroscience study using magnetic imaging, Reutskaja et al. (2018) document an inverted V-shaped relationship between brain activity and cognitive load, showing a turning point where processing new information outweighs the benefits. Thus, the task performance of a decision maker initially improves as more information is received, but beyond a certain threshold, additional information diminishes the quality of the cognitive processing. We investigate the implications of this neurological turning point for the financial markets, where participants are continually faced with and have to process new information.

In particular, we investigate the stock market response to information overload. 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] Using these novel data, we uncover a non-linear relationship between information flow and the aggregate risk premium: although receiving new information reduces the market risk premium, when investors face information overload, i.e., when the

[1] Excess information flow from mass and social media is not only a salient feature of the modern information age (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."

number of news articles exceeds a *dynamic* threshold, investors require a higher market risk premium.

Why does information overload affect the risk premium? Investors learn from news, however, they have a limited processing capacity. Thus, in periods of excessive information load, investors become cognitively overloaded and they will *not be able to* analyze some useful information.[2] Therefore, in such periods, extracted information is less precise than otherwise, implying that the estimated parameters of the true asset value are more likely to be wrong (see, for example, Coles and Loewenstein, 1988; Coles et al., 1995). Consequently, the risk-averse investor will require a higher risk premium to hold the asset because of increased estimation risk.

To study the relationship between information load and excess market returns, we calculate the number of financial market-related news articles in a given day (*InfLoad*) and the excessive information load (*InfOver*) when *InfLoad* deviates from a dynamic threshold, defined as its one standard deviation band calculated using twelve–month moving windows. By using a dynamic threshold, we enable information capacity to adapt to the changes in the number of news, including receiving more news during high uncertainty periods. This dynamic threshold also allows investors to digest more or less information as their learning abilities and resources can change over time. We find that a one-standard-deviation increase in *InfOver* increases the monthly market risk premium by about 60 basis points. *InfOver* contains additional explanatory power regarding market returns beyond the standard predictors of returns, macroeconomic factors, and other news-based measures. Moreover, when considering both the quantity and quality of the news, the

[2] While institutional investors have access to large teams, fast computers, and sophisticated tools, the sheer volume of information—news, market reports, earnings updates, economic indicators, etc.—can still overwhelm them. Even assuming an infinite processing capacity for institutional investors, today a non-negligible part of the trading (about 25% of the total trading volume of the U.S. equity markets, for example) is still done by the retail investors and thus information overload could still affect returns.

quantity matters the most when predicting market returns. This finding suggests that risk premia increase not due to investors accessing lower-quality information, but because they are overwhelmed by excessive information.

Our empirical analysis comes with two main challenges: (1) the supply of news is endogenous; and (2) the printed edition of *The New York Times* may not be a representative sample of news flows. First, the news supply is potentially correlated with fundamental shocks and increases in market-wide uncertainty. Consequently, it could be macroeconomic conditions and uncertainty that drive the market risk premium, rather than information overload. Similarly, unexpected stock return movements may lead to increased media coverage, complicating the direction of causality in our identification.

We attempt to address the first challenge via two approaches. First, we start with a diff-in-diff strategy, leveraging exogenous changes in the supply of news, including modifications in newspaper column formats. Second, we use non-market-related news (such as news on obituaries, fashion, and gaming) and generate instruments for information (over)load. Here, our identification assumption is that these alternative news items are unlikely to directly affect the stock market index and their supply is not driven by macroeconomic conditions. Newspapers may choose to feature financial markets related news on days with high uncertainty. However, we find a strong *positive* association between the number of non-market-related and financial market-related news items, supporting our view that the number of news changes based on editorial choices or the structure of the newspaper at the time, rather than financial market news crowding out other news during stress periods. Our main findings continue to hold under either approach, boosting our confidence that information overload increases excess stock returns in the next period.

The second identification challenge we face is that information overload is measured by relying only on the printed edition of a single newspaper, which undoubtedly does not comprehensively capture the entire flow of information agents face. To address this challenge, we demonstrate the relationship between information overload and stock market returns within a distinct setting, separate from the news flows originating from the media. Specifically, we measure information overload through the wave of earnings announcements made by the universe of U.S. publicly traded firms. We find 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. Consequently, the average contemporaneous stock price on days with information overload is lower than on days when fewer firms announce positive earnings estimates.

Furthermore, considering additional sources to quantify information overload would only exacerbate its effects on stock markets and thus, the sample selection bias likely works against us. Indeed, in a subsample analysis, we find that the effects of *InfOver*—measured via *The New York Times* only—are the strongest in the early period (before the 1930s), when newspapers were the main media outlets and *The New York Times* was one of the largest newspapers based on print circulation. As we move into the later decades–the 1950s, 1980s, and 1990s–this relationship weakens almost monotonically with the rise of radio, television, and the emergence of new information sources like Bloomberg and the internet.[3]

In our final set of results, we show that information overload on financial market related news does not affect all stock returns uniformly but has cross-sectional effects. Specifically, the impact of *InfOver* is stronger for smaller and more volatile stocks, and for stocks with a higher share of retail investors and those not included in the Standard and Poor's S&P500 index, compared to their peers. These findings suggest that a plethora of market related information makes valuation even more difficult and exacerbates estimation risk for stocks that require larger learning

[3]We still have the results from post-2000 and post-2010, when high-frequency trading was done intensively, demonstrating that our findings hold even in the presence of high-frequency traders.

efforts. Such results are in line with Akbas et al. (2018), who show that information processing is more difficult for stocks with lower institutional ownership, with Hong and Sraer (2016); Baker and Wurgler (2006) who argue that small and riskier stocks are more difficult to arbitrage, and with Bali et al. (2018) who show that unusual news predicts the cross-section of returns by driving idiosyncratic volatility. Moreover, we find evidence supporting 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.

This paper is related to three strands of literature. First, we contribute to the growing literature that links neuroscience to trading performance and financial decision making (Mazzonna and Peracchi, 2024; Karolyi et al., 2024; Grinblatt et al., 2012, 2011; Kuhnen and Knutson, 2011). Second, it builds upon an extensive body of literature from various disciplines, including 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; Lee, 2012). Here, using a large sample of historical data, we link cognitive ability and information overload to stock market returns. Third, the anomalous return in periods of information overload can be explained by agents' limited attention, as in Van Nieuwerburgh and Veldkamp (2010); Mondria (2010); DellaVigna and Pollet (2009); Hirshleifer et al. (2009); Da et al. (2011). This literature studies the effects of limited attention on information acquisition and individual asset returns. However, there is less direct evidence that limited attention affects macro outcomes, including the aggregate equity risk premium, which is the main focus of this paper. We also contribute to the aforementioned literature by documenting a non-linear relationship between information flow and asset returns.

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), who 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, similarly to the delay in the impounding of information into asset prices that is due to the complexity of information in Cohen and Lou (2012). Our results 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. In section 3, we first describe our econometric methodology, introduce the variables used in the regressions, and then present the results. Finally, we conclude in section 4.

## 2 Measuring information load

We construct the information load index in four steps. First, we scan the *full* content of the daily *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.[4]

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

[4]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`

regression machine learning tool.[5] 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 end 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 *markets* topic.

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 on day $d$, we assign a weight corresponding to the markets topic ($Mkt$):

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

where $C_{Mkt,j,d}$ and $C_{j,d}$ are the number of words associated with the markets topic and the total number of words appearing in the lead paragraph of article $j$ on day $d$, respectively. Thus, for each article and 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 $t$:

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

[5]We apply neural networks, the gradient method, and logistic regression algorithms. After training, testing, and validating the algorithms, we conclude that 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.

## 2.1 Descriptive analysis — information load

Figure 1 plots *InfLoad*, averaged across the 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 the 1980s and decreasing again after that. In our empirical analysis, we normalize the number of financial market news articles 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 the historical trend when calculating information overload.

The number of financial market news articles remained broadly unchanged until the early 1900s, followed by an increasing trend following the end of World War I. The 1920s, also known as the "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 articles also soared during that period and then started to decrease afterwards. Moreover, by 1930, about 40% 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 the mid-1980s, the first digital production of *The New York Times* started and after the mid-1990s the usage of the internet and other news outlets such as Bloomberg and social media changed the use of newspapers. Yet, upward trends can be observed even in recent periods, such as during the 2008 Global Financial Crisis.

Furthermore, in Table 1, we present contemporaneous Pearson correlation coeffi-

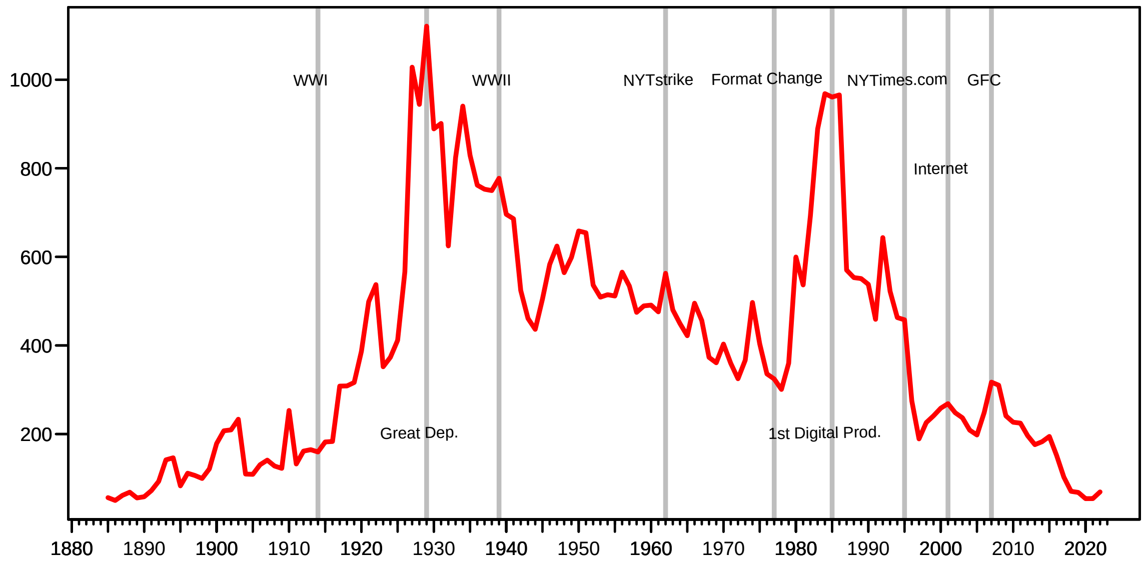


**Figure 1: Information Load**

The figure presents the annual averages of the total number of financial market-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.

cients between information load and 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). As proxies for 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, 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 a higher 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). As a proxy for 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 deviations 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 condition expectations and the standard deviation of price level expectations for the next 12 months. We document a positive correlation between the heterogeneity among households regarding 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 the St Louis Fed Financial Stress Index (STLFSI4) based on FRED economic data.

# 3 Empirical analysis

## 3.1 Econometric model

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

$$rx^m_{t+1} = \alpha_1 + \alpha_2 InfOver_t(tr) + \alpha_3 InfLoad_t + \Gamma Controls_t + \varepsilon_{t+1}, \tag{3}$$

where $rx^m_{t+1}$ is the one-period-ahead market return 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 dynamic threshold $tr_t$. We consider the one-standard-deviation band as the threshold, calculated using a twelve-month moving window following Danielsson et al. (2018, 2023) . One can think of this threshold as the maximum number of news articles that can be processed by the investor without her experiencing information overload. Specifically:

$$InfOver_t(tr) = \begin{cases} InfLoad_t - tr_t & \text{if} \quad InfLoad_t \geq tr_t \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The coefficient of interest is $\alpha_2$. We hypothesize a non-linear relationship between information flow and the market risk premium: 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 ($\Delta\text{SENT}_t$) as it is 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&P500 monthly dividend yield ($\mathrm{DY}_t$), following Shiller (1978); Campbell (1987); Fama and French (1988), among others. We obtain data from the Global Financial Data, Inc., GFDatabase (GFD). Second, we include realized volatility ($\mathrm{RVOLA}_t$), 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 traded volume ($\mathrm{VOL}_t$) as a proxy for market liquidity, defined as the total value of shares traded over a month, using price and volume data from CRSP. We also consider the default spread ($\mathrm{DS}_t$), the term spread ($\mathrm{TS}_t$), and the change in short-term interest rates ($\Delta\mathrm{STIR}_t$) following Keim and Stambaugh (1986); Campbell (1987); Fama and French (1989). $\mathrm{DS}_t$ is measured as the difference between Moody's Seasoned BAA and AAA corporate bond yields and the data are retrieved from FRED, the Federal Reserve Bank of St Louis (DBAA and DAAA, respectively). $\mathrm{TS}_t$ is calculated as the difference between the 10-year Treasury bond and the 3-month Treasury bill yields, and $\Delta\mathrm{STIR}_t$ is the change in the 3-month Treasury bill rate. We obtain interest rate data from the GFD. Finally, we include NBER recession dates to control for the business cycle as news flows are expected to increase during periods of uncertainty.

In Table 2, we present summary statistics. On average, about 500 financial market-related articles are printed at *The New York Times*, in a month, reaching a maximum of over 1,000, making up roughly one-third of the printed business articles. *InfOver* is correlated by about 47% with *InfLoad*, and by 18% and 11% with $\mathrm{DS}_t$ and $\mathrm{DY}_t$, respectively, suggesting that information overload is not likely to share common information with standard predictors of stock market returns. The Phillips-Perron stationarity tests strongly reject the null of the unit root for all of the variables.

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

Table 3 presents the estimated coefficients from (3). 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 information load has a non-linear relationship with stock market returns.

To incorporate such non-linearity, 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 about 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. The estimated coefficients of *InfOver* and *InfLoad* are statistically different from each other.

In column III, we further consider the "quality" of the news in addition to the quantity and include consistency in the tone of the article, calculated as squared sentiment (SENT2). The higher is SENT2, the more consistent are the articles on average because the tone is becoming more and more optimistic or pessimistic. A priori, investors are more likely to be confused when news becomes inconsistent 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.

Finally, we test the explanatory power of information overload for future market returns compared with other text-based measures: Sentiment (SENT), the Economic Policy Uncertainty index (EPU) of Baker et al. (2016), the Geopolitical

Risk Index, GPR, of Caldara and Iacoviello (2022), and News Implied Volatility (NVIX) of Manela and Moreira (2017). Columns IV and V show that *InfOver* has about 1% incremental explanatory power for 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 to be expected, given that both EPU and GPR are shown to be useful predictors of macroeconomic series rather than financial series.

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 60% of the variation in the stock market returns. Yet, information load and overload explain about 30% of the variation, whereas the other news-based measures together explain just above 10% of the variation, with the contribution of EPU higher than that of GPR and NVIX.

## 3.3 Effects of information overload on stock returns: Endogeneity concerns

So far, we have shown that information overload predicts stock market returns. However, such inferences 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 articles. Furthermore, omitted variables could affect both information load and market returns, simultaneously. For example, worsening economic conditions increase the supply of information and also the 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 the home as an instrument.[6] We choose these non-market-related topics because they are very unlikely to directly affect stock market returns, but their information load is likely to be correlated with that of the financial market news, with 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.

To generate the instrumental variables (IVs), we follow our main methodology for calculating information load and overload using this alternative set of news articles instead of the financial market news, denoting the outcomes by $\textit{InfLoad}^{IV}$ and $\textit{InfOver}^{IV}$. 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 $\textit{InfOver}^{IV}$ and *InfOver* and $\textit{InfLoad}^{IV}$ and *InfLoad*. The documented positive relationship between the number of financial market and non-market-related news articles is crucial because one could argue that these IVs violate the exclusion restriction as newspapers may choose to feature more non-market-related news on quiet market days. However, we find that financial market-related news does not crowd out non-market-related news. Instead, the two move hand-in-hand. Furthermore, the $F$-statistics are 23 and 67, respectively, both significantly greater than 10, supporting the relevance of the IVs. The second-stage regression (column III) provides supporting evidence on the causal effect of information overload on the next period's stock market returns.

As a second approach to addressing 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

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

published in *The New York Times*. For example, we include newspaper strikes, changes in the physical size of the print edition, editorial decisions regarding the combination of some sections in the newspaper, and when `NYTimes.com` began publishing, and run the following regression:[7]

$$\begin{aligned} rx^{m}_{t+1} &= \alpha_1 + \alpha_2 \mathit{InfOver}_t(tr) + \alpha_3 \mathit{InfOver}_t(tr) \times D_{i,t} + \alpha_4 D_{i,t} \\ &+ \alpha_5 \mathit{InfLoad}_t + Controls + \varepsilon_{t+1}, \end{aligned} \tag{5}$$

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.

## 3.4 Effects of information overload on stock returns: Sample selection concerns

We measure information overload by relying solely on the printed edition of a single newspaper. We acknowledge that data from *The New York Times* do not

[7]The strikes could depend on the profitability of the newspaper at the time, which in turn is a function of macroeconomic conditions and thus may not be exogenous. However, the 1978 New York City newspaper strike, for instance, halted the production of the three major newspapers: The New York Times, the New York Daily News, and the New York Post. Unlike in many strikes, wage levels themselves were not the issue; instead, the newly issued work rulings (such as the relaxed requirements concerning the level of staffing) drove the strikes. See "The Times and News resume publication", November 6, 1978, *The New York Times.*

comprehensively capture the entire flow of information. Undoubtedly, investors can learn from other newspapers, the internet, or other sources such as Reuters and Bloomberg. To address this issue, we employ two approaches.

First, we examine the relationship between information overload and market returns in different subsamples. We cover a very long time series from July 1926 to December 2022, which is subject to significant changes both in terms of the stock market and information flow dynamics. We run the baseline specification (3) using a sample period from the beginning of the total period up to the 1930s, up to the 1940s, and then moving forward in ten-year increments up to the 2020s. Figure 2 plots the estimated coefficients of *InfOver* for each time periods considered. The impact of information overload on stock market returns was most pronounced in the early period (before the 1930s), when newspapers were the primary media outlets and *The New York Times* was among the largest based on print circulation. The relationship decreases almost monotonically after the 1950s, in the 1980s, and in the 1990s, with the wide usage of radio, TV, and the foundation of other outlets such as Bloomberg and the internet, respectively. Consequently, we argue that our sample selection bias likely underestimates the true extent of information overload, and incorporating additional sources would only amplify its effects on stock market returns.

In our second approach, we aim to confirm our main finding in a different setting, other than measuring information overload through the number of articles in a newspaper. Specifically, we use clustered earnings announcements to quantify 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

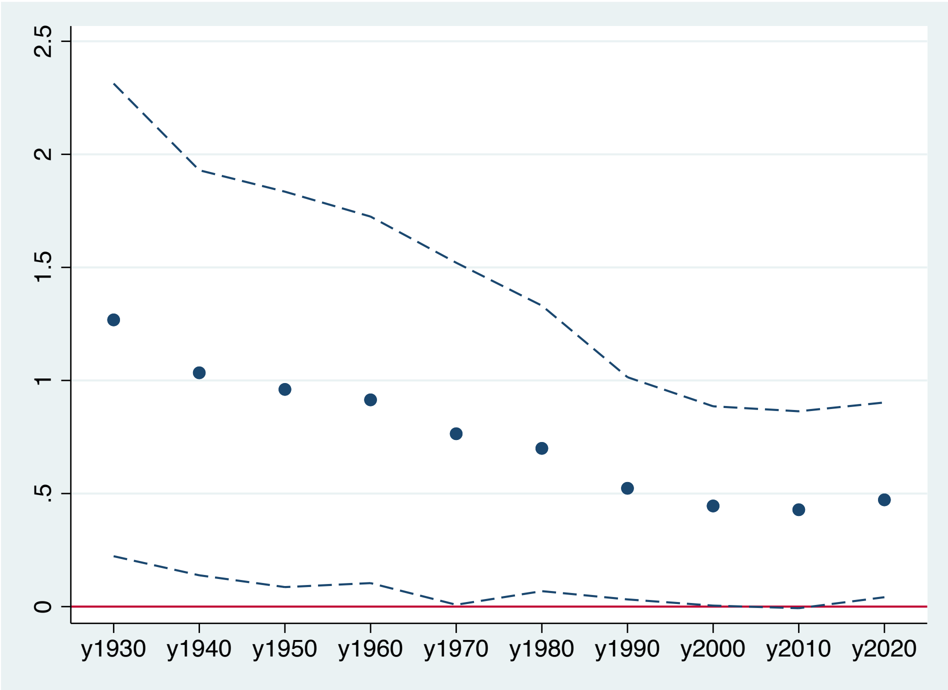


**Figure 2: The effects of information overload on stock market returns in different periods**

The figure presents the estimated standardized coefficients of *InfOver* from running regressions (3) for different time periods considered labeled in the x-axis. Specifically, "y1930" corresponds to a sample period which includes data from 1885 until 1930, "y1990" spans data from 1885 until 1990, for example.We also plot the 95% confidence intervals bands for the estimated coefficients.

(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. Consequently, 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 on the same day—information overload— it becomes challenging for investors to focus on each announcement adequately. Thus, the effects of earnings announcements on returns should depend on the number of firms making these announcements.

To test our hypothesis, we collect from I/B/E/S earnings announcement data provided by 15,463 firms, spanning from January 1985 to December 2022. We focus only on positive announcements when the actual earnings exceed the average earnings per share forecasts for a firm in a given forecast period. We then split the data into quintiles depending on the number of firms announcing their earnings on the same day.[8] Specifically, Group 1 corresponds to the lowest quintile, i.e., the days with the smallest numbers of firms making announcements, while Group 5 corresponds to the highest quintile. We then create five dummy variables, $D_t^j$, equal to 1 if the number of earnings announcements on day $t$ falls into Group $j$, and run the following regression:

$$r_{i,t} \;=\; \sum_{j=1}^{5} \alpha_j D_t^j + \beta r x_t^m + \gamma esurp_{i,t}|_{esurp_{i,t}>0} + \varepsilon_{i,t}, \tag{6}$$

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 earnings cluster group, the average returns are positive on the days with 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 numbers of firms making announcements) than for Group 5 (the bucket representing the days with the largest number of firms making announcements), all else equal. Thus, on the days in which earnings announcements cluster—investors face information overload—the average returns are significantly lower than on the days in which fewer firms are making announcements.

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

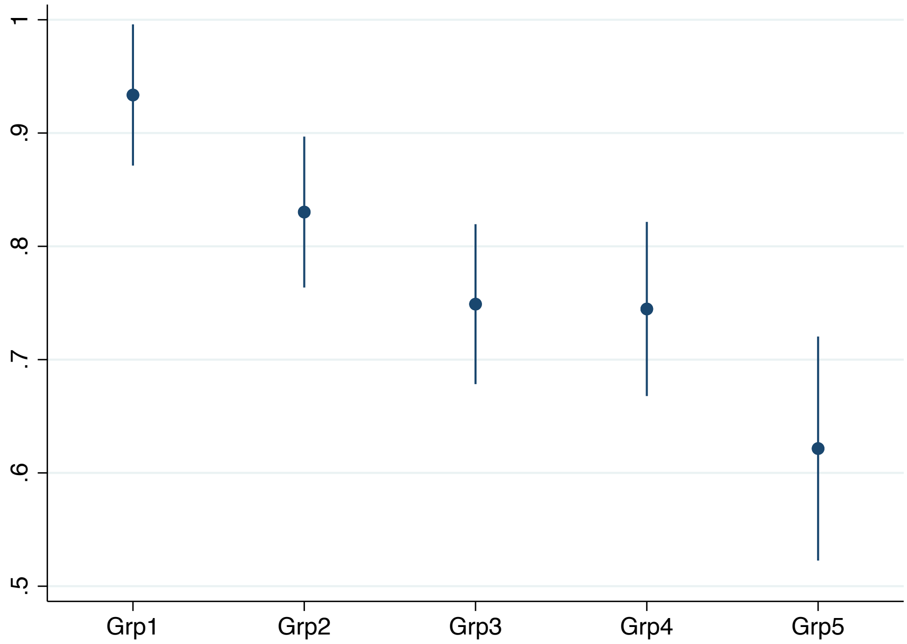


**Figure 3: The effects of clustered earnings on stock returns**

The figure presents the estimated coefficients of groups ($\widehat{\alpha}$s) in (6) along with their 95% confidence intervals. Group 1 corresponds to the lowest quintile, i.e., considers the days with the smallest numbers 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.

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

We expect information overload on market related news to have cross-sectional effects on individual stock returns, influenced by firm characteristics. This is because, in periods of information overload, investors are cognitively constrained and it is optimal for them to allocate their processing capacities into a certain group of stocks that require 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 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 from the S&P500 index and stocks with lower institutional ownership as retail investors have fewer resources than 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 will require a higher risk premium to hold smaller and more volatile stocks, stocks that are not included in the S&P500 index, and those with lower institutional ownership. To examine these conjectures, we run the following regression model:

$$\begin{aligned} r_{i,t+1} &= \alpha_1 + \alpha_2 \mathit{InfOver}_t(tr) \times D_{i,t}^{\mathit{Grp1}} + \alpha_3 \mathit{InfOver}_t(tr) \times D_{i,t}^{\mathit{Grp2}} \\ &+ \alpha_4 \mathit{InfLoad}_t + \mathit{Controls} + \varepsilon_{t+h}, \end{aligned} \tag{7}$$

where $r_{i,t+1}$ is the stock return of firm $i$ in month $t+1$. $D_{i,t}^{\mathit{Grp1}}$ is a dummy variable that equals 1 if, in month $t$, stock $i$ is small and highly volatile, is not a constituent of the S&P500 index, and has a smaller share of institutional ownership of the shares outstanding. Analogously, $D_{i,t}^{\mathit{Grp2}}$ is a dummy variable that equals 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 top and bottom 20% 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 (7). Column I shows that, unconditionally, information overload not only leads to higher stock market returns (as documented so far), but also higher individual stock returns. Columns II-VI show that smaller and more volatile stocks, stocks that are not included in the S&P500 index, and those with lower institutional ownership, exhibit significantly higher returns than their counterparts during periods of information overload. Thus, the effects of information overload on next-period returns depend on firm characteristics and that information overload exacerbates mispricing by aggravating investors' capacity constraints.

## 3.6 Robustness

To test the sensitivity of our findings, we run several robustness tests. We start by altering the definition of information overload by using either different thresholds or different estimation windows. First, instead of calculating information overload using the deviation over a one-standard-deviation band, we measure it via the deviations from alternative thresholds: the historical mean and trend estimated via the Hamilton (2017) filter. Second, instead of using a 12-month rolling window to calculate the one-standard-deviation threshold, we employ moving window sizes of 3, 6, and 24 months.

Third, instead of calculating stock market volatility as the standard deviation of daily stock returns in a given month, we estimate it via a GARCH(1,1) model (3).

Fourth, we set the maximum lag order of autocorrelation to 6 instead of 2 when calculating the Newey-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 change in the U.S. consumer price index, the change in the industrial production index, and the 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 periods in which the S&P500 index decreases by over 10%. In addition, we exclude major global episodes: the Great Depression (1929–1935), the Global Financial Crisis (2007–2009), and the 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 next-period excess market returns.

# 4 Conclusion

In this paper, we study the role of media news in stock markets and provide the first evidence that an excessive amount of information predicts stock returns. 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, which can be thought of as the representative agent's processing capacity limit. We find that information flow is valuable and agents learn from news. However, the relationship between information flow and stock returns is non-linear. Following excessive news-flow periods, investors require a higher risk premium in return for holding 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.

# 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.

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.

Bali, T. G., A. Bodnaruk, A. Scherbina, and Y. Tang (2018). Unusual news flow and the cross section of stock returns. *Management Science 64*(9), 4137–4155.

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 (2018). Learning from history: Volatility and financial crises. *The Review of Financial Studies 31*(7), 2774–2805.

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.

Grinblatt, M., M. Keloharju, and J. Linnainmaa (2011). Iq and stock market participation. *The Journal of Finance 66*(6), 2121–2164.

Grinblatt, M., M. Keloharju, and J. T. Linnainmaa (2012). Iq, trading behavior, and performance. *Journal of Financial Economics 104*(2), 339–362.

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.

Karolyi, S., T. Ruchti, and P. Truong (2024). Number processing constraints and earnings news. *Management Science*.

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

Kuhnen, C. M. and B. Knutson (2011). The influence of affect on beliefs, preferences, and financial decisions. *Journal of Financial and Quantitative Analysis 46*(3), 605–626.

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.

Mazzonna, F. and F. Peracchi (2024). Are older people aware of their cognitive decline? misperception and financial decision-making. *Journal of Political Economy 132*(6), 000–000.

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

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory 145*(5), 1837–1864.

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*.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and under-diversification. *The Review of Economic Studies 77*(2), 779–805.

**Table 1: Pearson correlation coefficients**

In this table, we present contemporaneous Pearson correlation coefficients ($\rho$) between information load and proxies for 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 the 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 condition expectations, respectively, obtained from Michigan Surveys of Consumers. Finally, NFCI is the Chicago Fed National Financial Conditions Index and STLFSI4 is the 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 |
| $\rho$ | 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 $p$-values corresponding to the Phillips-Perron stationarity test results for the main variables included in our analysis. The last row reports the series' correlation between the column header and information overload $InfOver_t$, which is introduced in (4). 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 market-related news articles in a given month. $\Delta \mathrm{SENT}_t$ is the first difference of the market sentiment measure, calculated as the difference between the proportions of positive and negative words, following Garcia (2013). $\mathrm{DY}_t$ is the S&P500 monthly dividend yield, $\Delta \mathrm{STIR}_t$ is the change in the three-month Treasury bill rate, $\mathrm{TS}_t$ is the term spread, calculated as the difference between the ten-year Treasury bond and the three-month Treasury bill yields. Default spread, $\mathrm{DS}_t$, is measured as the difference between the BAA and AAA corporate bond spreads, $\mathrm{RVOLA}_t$ is the monthly realized volatility, calculated as the standard deviation of stock market returns. Finally, $\mathrm{VOL}_t$ is the total value of shares traded over a month. 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, the Center for Research in Security Prices (CRSP) 1925 US Indices Database, and Wharton Research Data Services.

| | I | II | III | IV | V | VI | VII | VIII | IX | X |
|---|---|---|---|---|---|---|---|---|---|---|
| | $rx_t^m$ | $InfOver_t$ | $InfLoad_t$ | $\Delta \mathrm{SENT}_t$ | $\mathrm{DY}_t$ | $\Delta \mathrm{STIR}_t$ | $\mathrm{TS}_t$ | $\mathrm{DS}_t$ | $\mathrm{RVOLA}_t$ | $\mathrm{VOL}_t$ |
| mean | 0.68 | 14.10 | 481.47 | -0.00 | 3.64 | 0.00 | 1.63 | 1.12 | 0.04 | 42.59 |
| min | -29.13 | 0.00 | 41.08 | -0.35 | 1.08 | -3.85 | -2.64 | 0.32 | 0.01 | 0.01 |
| max | 38.85 | 470.54 | 1154.33 | 0.36 | 9.63 | 2.60 | 4.41 | 5.64 | 0.27 | 753.53 |
| stdev | 5.34 | 53.83 | 269.05 | 0.07 | 1.61 | 0.41 | 1.26 | 0.68 | 0.03 | 98.97 |
| PP–stationarity | 0.01 | 0.01 | 0.01 | 0.01 | 0.03 | 0.01 | 0.01 | 0.01 | 0.01 | 0.09 |
| corr(InfOver) | -0.02 | 1.00 | 0.47 | 0.04 | 0.11 | 0.03 | -0.05 | 0.18 | 0.08 | -0.08 |

**Table 3: Information overload and market excess returns**

In this table, we report the estimated coefficients of the time-series regressions introduced in (3). 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 (4), respectively. SENT2 is the squared sentiment measure introduced in Section 3.1. 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 at a monthly frequency from July 1926 to December 2022, where available, and standardized to ease the interpretation of the coefficients. Newey-West standard errors with 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, the Center for Research in Security Prices (CRSP) 1925 US Indices Database, and Wharton Research Data Services.

| Dep. var.: $rx_{t+1}^m$ | I | II | III | IV | V |
|---|---|---|---|---|---|
| $InfOver_t$ | | 0.63*** | 0.63*** | | 0.66** |
| | | (0.237) | (0.241) | | (0.257) |
| $InfLoad_t$ | -0.11 | -0.58* | -0.60* | | -0.63* |
| | (0.284) | (0.314) | (0.356) | | (0.327) |
| $\text{SENT}_t^2$ | | | -0.04 | | |
| | | | (0.188) | | |
| $\Delta\text{SENT}_t$ | -0.12 | -0.13 | -0.14 | 0.02 | 0.03 |
| | (0.197) | (0.196) | (0.200) | (0.229) | (0.230) |
| $\text{EPU}_t$ | | | | 0.22 | 0.19 |
| | | | | (0.281) | (0.279) |
| $\text{GPR}_t$ | | | | 0.06 | 0.03 |
| | | | | (0.190) | (0.211) |
| $\text{NVIX}_t$ | | | | -0.00 | 0.09 |
| | | | | (0.222) | (0.205) |
| $\text{DY}_t$ | 0.46* | 0.61** | 0.61** | 0.34 | 0.47 |
| | (0.238) | (0.240) | (0.238) | (0.266) | (0.294) |
| $\Delta\text{STIR}_t$ | -0.34 | -0.37 | -0.37 | -0.32 | -0.34 |
| | (0.243) | (0.231) | (0.231) | (0.247) | (0.235) |
| $\text{TS}_t$ | 0.14 | 0.19 | 0.18 | 0.11 | 0.19 |
| | (0.181) | (0.179) | (0.180) | (0.182) | (0.184) |
| $\text{DS}_t$ | 0.30 | 0.25 | 0.25 | 0.27 | 0.24 |
| | (0.533) | (0.532) | (0.542) | (0.560) | (0.576) |
| $\text{RVOLA}_t$ | 0.01 | 0.05 | 0.05 | -0.15 | -0.12 |
| | (0.363) | (0.355) | (0.359) | (0.431) | (0.420) |
| $\text{VOL}_t$ | 0.12 | -0.02 | -0.02 | -0.21 | -0.65 |
| | (0.219) | (0.224) | (0.232) | (0.612) | (0.577) |
| NBER | -1.59*** | -1.80*** | -1.79*** | -1.39** | -1.63*** |
| | (0.566) | (0.588) | (0.585) | (0.571) | (0.596) |
| Adj. $R^2$ (%) | 1.26 | 2.53 | 2.45 | 1.18 | 2.46 |
| No Obs. | 1,150 | 1,150 | 1150 | 1055 | 1055 |
| | | | | | |
| $p-$values | | | | | |
| $H_0 : b_{InfOver} = b_{InfLoad}$ | | <0.001 | <0.001 | | <0.001 |

**Table 4: Information overload and market excess returns–Endogeneity concerns** *InfLoad* and *InfOver* are introduced in (2) and (4), respectively. $InfLoad^{IV}$ and $InfOver^{IV}$ are instruments for *InfLoad* and *InfOver*, calculated using articles related to obituaries, fashion, gaming, health, arts, travel, and the home instead of financial market news. Columns I and II report the first stage of the two-stage least squares (2SLS) regressions for both instrumental variables. Column III reports the second-stage regression results. In columns IV and V, we report the estimated coefficients from (5), 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 at a 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, the Center for Research in Security Prices (CRSP) 1925 US Indices Database, and Wharton Research Data Services.

| | I | II | III | IV | V |
|---|---|---|---|---|---|
| | 1st stage | 1st stage | 2nd stage | exo. events | exo. events |
| | *InfOver* | *InfLoad* | | | |
| $InfOver_t$ | | | 1.34** | 0.63*** | 0.63*** |
| | | | (0.533) | (0.237) | (0.237) |
| $InfOver^{IV}$ | 0.34*** | 0.05** | | | |
| | (0.029) | (0.020) | | | |
| $InfLoad_t$ | | | -1.90 | -0.58* | -0.58* |
| | | | (1.366) | (0.315) | (0.314) |
| $InfLoad^{IV}$ | 0.08** | 0.16*** | | | |
| | (0.038) | ( 0.026) | | | |
| $InfOver_t(\tau) \times D_{i,t}$ | | | | -0.10** | -0.11*** |
| | | | | (0.045) | (0.026) |
| $D_{i,t}$ | | | | -0.25 | |
| | | | | (1.182) | |
| $\mathrm{SENT}_t$ | 0.04 | 0.04* | -0.10 | -0.13 | -0.13 |
| | (0.038) | (0.026) | (0.194) | (0.196) | (0.196) |
| $\mathrm{DY}_t$ | -0.07 | 0.19*** | 1.02** | 0.61** | 0.61** |
| | (0.047) | (0.032) | (0.458) | (0.239) | (0.240) |
| $\Delta\mathrm{STIR}_t$ | 0.02 | 0.002 | -0.39** | -0.37 | -0.37 |
| | (0.032) | (0.022) | (0.160) | (0.232) | (0.232) |
| $\mathrm{TS}_t$ | -0.07** | 0.004 | 0.22 | 0.19 | 0.19 |
| | (0.036) | (0.025) | (0.178) | (0.179) | (0.179) |
| $\mathrm{DS}_t$ | 0.18*** | 0.18*** | 0.31 | 0.25 | 0.24 |
| | (0.046) | (0.032) | (0.286) | (0.533) | (0.532) |
| $\mathrm{RVOLA}_t$ | -0.02 | 0.04 | 0.14 | 0.05 | 0.05 |
| | (0.041) | (0.029) | (0.222) | (0.356) | (0.355) |
| $\mathrm{VOL}_t$ | -0.05 | -0.33*** | -0.46 | -0.01 | -0.01 |
| | (0.038) | (0.026) | (0.515) | (0.226) | (0.225) |
| NBER | 0.27*** | -0.02 | -2.05*** | -1.78*** | -1.78*** |
| | (0.093) | (0.064) | (0.498) | (0.588) | (0.589) |

**Table 5: Information overload and cross-section of returns**
In this table, we report the estimated coefficients from (7). The dependent variable is $r_{i,t+1}$, the stock returns of firm $i$ in month $t+1$. *InfLoad* and *InfOver* are introduced in (2) and (4), respectively. $D^{gr1}_{i,t}$ is a dummy variable that equals 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 of shares outstanding. Analogously, $D^{gr2}_{i,t}$ is a dummy variable that equals 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 top and bottom 20% of the corresponding firm characteristics. The rest of the variables are described in Table 2. All of the variables are at a 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, the Center for Research in Security Prices (CRSP) 1925 US Indices Database, Wharton Research Data Services, and the Thomson Reuters Stock Ownership database.

| Dep. var.: $r_{i,t+1}$ | I<br>baseline | II<br>size | III<br>volatility | IV<br>S&P500 | V<br>inst. ownership |
|---|---|---|---|---|---|
| $InfOver_t$ | 0.52*** | | | | |
| | (0.191) | | | | |
| $InfOver_t \times D^{gr1}_{i,t}$ | | 0.50*** | 0.36*** | 0.25** | 0.19** |
| | | (0.187) | (0.102) | (0.111) | (0.082) |
| $InfOver_t \times D^{gr2}_{i,t}$ | | 0.14*** | 0.13** | 0.13*** | 0.10 |
| | | (0.052) | (0.060) | (0.047) | (0.077) |
| $InfLoad_t$ | -0.67*** | -0.67*** | -0.49** | -0.47** | -0.26 |
| | (0.236) | (0.236) | (0.226) | (0.225) | (0.305) |
| $\text{SENT}_t$ | -0.05 | -0.05 | -0.04 | -0.03 | -0.16 |
| | (0.187) | (0.187) | (0.187) | (0.188) | (0.219) |
| $\text{DY}_t$ | 0.67** | 0.67** | 0.63** | 0.62** | 0.28 |
| | (0.275) | (0.275) | (0.275) | (0.275) | (0.640) |
| $\Delta\text{STIR}_t$ | -0.54*** | -0.54*** | -0.53*** | -0.53*** | -0.37 |
| | (0.190) | (0.190) | (0.191) | (0.192) | (0.231) |
| $\text{TS}_t$ | 0.17 | 0.17 | 0.14 | 0.14 | 0.12 |
| | (0.216) | (0.216) | (0.215) | (0.215) | (0.246) |
| $\text{DS}_t$ | 0.84** | 0.84** | 0.85*** | 0.86*** | 0.90 |
| | (0.329) | (0.329) | (0.329) | (0.330) | (0.617) |
| $\text{RVOLA}_t$ | 0.06 | 0.06 | 0.05 | 0.05 | 0.13 |
| | (0.415) | (0.415) | (0.417) | (0.418) | (0.520) |
| $\text{VOL}_t$ | -0.23 | -0.23 | -0.20 | -0.19 | -0.19 |
| | (0.312) | (0.312) | (0.312) | (0.312) | (0.325) |
| NBER | -2.38*** | -2.37*** | -2.26*** | -2.25*** | -2.13 |
| | (0.845) | (0.845) | (0.842) | (0.842) | (1.358) |
| Adj. $R^2$ (%) | 0.79 | 0.79 | 0.76 | 0.74 | 0.95 |
| No Obs. | 4,666,461 | 4,666,461 | 4,662,778 | 4,665,535 | 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 III, we define the threshold as the historical mean and the trend estimated via the Hamilton (2017) filter. In columns IV to VI, we set 3, 6, and 24-month rolling windows to calculate the 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 the industrial production index, CPI inflation, and the DLR of Danielsson et al. (2023) as a proxy for risk appetite). In column X, we exclude major stock market crashes, defined as the S&P500 index dropping by over 10%. Finally, in column XI, we exclude major global episodes: the Great Depression (1929–1935), the Global Financial Crisis (2007–2009), and the World Wars (1914–1918, 1939–1945). $InfOver_t$ is the information overload measure introduced in (4). 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 at a monthly frequency from July 1926 to December 2022, where available, and standardized to ease the interpretation of the coefficients. Newey-West standard errors with a 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, the Center for Research in Security Prices (CRSP) 1925 US Indices Database, and 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 |
| $InfOver_t$ | 0.63*** | 0.86*** | 0.72*** | 0.52*** | 0.55** | 0.57*** | 0.62** | 0.63*** | 0.64*** | 0.55*** | 0.31** |
| | (0.237) | (0.286) | (0.216) | (0.197) | (0.213) | (0.187) | (0.257) | (0.223) | (0.246) | (0.194) | (0.156) |
| $InfLoad_t$ | -0.58* | -1.02** | -0.85** | -0.46 | -0.49 | -0.58* | -0.60* | -0.58** | -0.70* | -0.59** | -0.15 |
| | (0.314) | (0.425) | (0.395) | (0.305) | (0.303) | (0.322) | (0.317) | (0.275) | (0.371) | (0.296) | (0.219) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adj. $R^2$ (%) | 2.53 | 2.89 | 2.33 | 2.19 | 2.29 | 2.18 | 2.63 | 2.53 | 2.68 | 2.78 | 2.16 |
| No Obs. | 1,150 | 1,150 | 1,150 | 1,150 | 1,150 | 1,150 | 1,140 | 1,150 | 1,082 | 1,118 | 946 |