Media Narratives and Price Informativeness*

Chukwuma Dim† **Francesco Sangiorgi**‡ **Grigory Vilkov**§

This version: July 27, 2023

Abstract

We show that an increase in stock return exposure to media attention to narratives, measured with standard methods for extracting topic attention from news text, leads to a lower stock price informativeness about future fundamentals. Empirically, narrative exposure explains over 86% of idiosyncratic variance in the cross-section, and both narrative exposure and non-systematic information channels—idiosyncratic variance and variance related to public information—decrease stock price informativeness. Moreover, stocks with high narrative exposure demonstrate elevated trading volume. To rationalize the empirical results, we suggest a mechanism based on disagreement among investors arising due to the differential processing of information in media narratives.

Keywords: media narratives, media bias, price informativeness, idiosyncratic risk, noise trading, latent demand.

JEL: G11, G12, G13, G17

*We received helpful comments and suggestions from Carol Alexander, Asaf Manela, Jonathan Brogaard, Monique Donders (discussant), Nandini Gupta, Yigitcan Karabulut, Badrinath Kottimukkalur, Emanuel Mönch, and Marina Niessner. We also thank seminar participants at the Frankfurt School of Finance & Management, George Washington University, Quantitative FinTech Group at Sussex, Rutgers School of Business – Camden, the University of Basel, and Empirical Asset Pricing Meeting in VU Amsterdam.

†Chukwuma Dim is with George Washington University, USA, `cdim@gwu.edu`.

‡Francesco Sangiorgi is with Frankfurt School of Finance & Management, Germany, `F.Sangiorgi@fs.de`.

§Grigory Vilkov is with Frankfurt School of Finance & Management, Germany, `vilkov@vilkov.net`.

1 Introduction

While slow-moving fundamentals play an essential role in asset pricing models as drivers of risks and risk premiums, in everyday lives, investors are constantly exposed to an intense flow of news containing informative, uninformative, and potentially biased signals from media outlets. Building on Robert Shiller's insights on the link between narratives and economic behavior (see, Shiller, 2020), a growing body of research now extracts narratives from the news and analyzes how media's attention to various narratives is related to economic quantities. Our main objective is to study how media attention to narratives affects the information embedded in stock prices in a typical environment, where differential interpretations of potentially biased news delivered by the media lead to disagreement among investors.[1] Do asset prices co-move with the intensity of coverage of specific narratives in the news media? Do stock prices that fluctuate stronger with narrative attention aggregate more information about future fundamentals? Does stock price exposure to media narrative attention create excess volatility? We address these and related questions empirically and suggest a plausible theoretical mechanism explaining the results.

First, we empirically show that individual stocks' price informativeness decreases for stocks that strongly co-move with media attention to narratives, with the effect more pronounced for smaller speculative stocks with lower institutional ownership. Second, we demonstrate that while adding narrative attention to standard factor models boosts a model's explanatory power by a tiny fraction (less than 0.1% adjusted R^2, on average), narrative exposure—defined as the weighted average intensity of return co-movement with individual narrative's attention—turns out to be the most prominent cross-sectional explanatory variable for idiosyncratic variance. The firm-specific public information component of return variance primarily drives the pattern, resulting in the level of the idiosyncratic and public-information-related variances being also linked to lower stock price informativeness. Third, we find that stocks strongly exposed to narrative attention experience higher turnover, which supports the role of media narrative exposure in explaining dispersion in latent demand across assets. Finally, we rationalize our findings with

[1] A growing body of research highlights news media biases and their implications for financial markets (e.g., Mullainathan and Shleifer, 2005; Baloria and Heese, 2018; Niessner and So, 2018; Goldman, Gupta, and Israelsen, 2021; Goldman, Martel, and Schneemeier, 2022) and an established literature studies departure from full rationality when agents process information and form beliefs (see Barberis, 2018 for a review). Cookson and Niessner (2020) document that investors' decisions are affected by both different information sets and differential interpretation of information.

a stylized dynamic trading model featuring media bias. The model shows that in an economy with biased media populated by some unsophisticated agents who cannot distinguish bias from true signals, asset returns correlate with *media attention* to narratives. In turn, the informativeness of asset prices diminishes with higher narrative exposures, ultimately leading to higher non-systematic variance in asset returns.

We use a large archive (more than 300,000) of online Wall Street Journal (WSJ) news articles to measure the news media's attention to narratives, which we extract for the period 1998 - 2021 using the Latent Dirichlet Allocation (LDA) algorithm. The procedure optimally identifies 33 narratives, covering issues related to politics, regulation, natural resources, fixed income, equity markets, and entertainment, among others. Importantly, media attention to each narrative varies substantially over time, reflecting, amongst other things, changing social and economic conditions, readers' interests, and media preferences, which have been shown to inhabit biases related to negativity (Gurun and Butler, 2012; Niessner and So, 2018), local bias (Gurun and Butler, 2012), political slant (Gentzkow and Shapiro, 2010), and political polarization (Goldman, Gupta, and Israelsen, 2021). We measure stocks' exposure to media narrative attention by regressing each stock's daily excess returns on standard factors augmented with the media's attention to a narrative, then take the latter's coefficient as a narrative beta. We define a stock's narrative exposure as the weighted average of absolute narrative betas for all identified narratives. Empirically, our proposed measure barely correlates with the number of mentions of a given firm in the news and, thus, is markedly different from standard stock-specific news coverage.

To test the link between narrative exposure and price informativeness, we adopt a microfounded stock-level measure of price informativeness based on Bai, Philippon, and Savov (2016), defined as the predicted variation in cash flows using current market prices. Bai, Philippon, and Savov (2016) demonstrate that this measure is also justified as a welfare measure using Q-theory.[2]

[2]In addition to its solid theoretical foundation and empirical support, we prefer this measure to the often-used nonsynchronicity measure—defined as $1 - R^2$ from a market model—because nonsynchronicity ambiguously captures both noise and potentially firm-specific information in stock prices. For example, a decrease (increase) in R^2 (nonsynchronicity) can be entirely due to noisy prices without any improvement in price informativeness, and vice versa. Accordingly, Brogaard, Nguyen, Putnins, and Wu (2022) show that despite the recent increases in R^2, which implies less informative prices based on nonsynchronicity, stock prices have instead increasingly reflected more firm-specific information. They also show that nonsynchronicity yields implausible relationships between price informativeness and several firm characteristics.

We employ a two-stage methodology where we first run an annual cross-sectional regression of future firm fundamentals on current market value and its interaction with narrative exposure and controls, then test whether the average coefficients differ from zero in the second stage. We observe lower price informativeness in stocks with higher narrative exposure, especially when the average market-wide narrative exposure is high. Importantly, *an increase* in a firm's narrative exposure leads to a decline in the informativeness of its price relative to those of similar firms that did not experience an increase in narrative exposure.

We establish the second major result by showing that narrative exposure alone explains over 86% of the cross-sectional variation in stocks' idiosyncratic risk. We use the Brogaard, Nguyen, Putnins, and Wu (2022) (BNPW) approach to decompose non-systematic variance into private and public firm-specific information and noise and find that narrative exposure is most closely related to the public information component, with noise and private information following closely behind. We find a consistent pattern between the different components of stock variance and price informativeness: high idiosyncratic and public-information-related variances are the strongest detractors of the information contained in stock prices. Finally, we provide consistent evidence for a positive link between shocks to narratives' attention and the turnover of stocks highly affected by such shocks.

To rationalize these results, we develop a stylized trading model with time-varying public information that addresses the following questions: Why would stock returns co-move with changes in news media's attention to narratives? How would exposure to narratives relate to price informativeness in the cross-section? The major driver behind the model is the bias in narratives delivered by media outlets. Following Mullainathan and Shleifer (2005), we acknowledge that media "bias is not a bug but a feature"[3] of the news media industry and that narratives can contain biases for a number of reasons, including to serve as a selling point for specific media outlets. To ensure close alignment with the current state of empirical research, the model maps the LDA algorithm—used in our empirical analysis—to the information process investors face.

An overview of the setup is as follows. A media outlet publishes news articles around several narratives correlated with firms' fundamentals. The amount of attention accorded to a narrative determines the number of articles on that narrative, and the narrative attention evolves randomly

[3] www.nytimes.com/2005/05/19/business/media/another-view-of-news-bias-as-selling-point.html

over time. Articles are informative but are also biased, and a fraction of investors do not account for this bias.[4] To derive a clear message about price informativeness, we assume investors are risk neutral and hence shut down any impact of narratives on risk premiums.

The model provides the following insights: (i) When attention to a narrative increases, the associated bias receives more weight in the unsophisticated investors' beliefs. Because asset prices reflect these beliefs, stock returns move in the direction of the narrative bias adjusted for cash flow narrative exposures. In this way, the model provides a mechanism for stock return covariance with changes in narrative attention—that is, for stocks' exposure to media narratives. (ii) The bias-related stock price reaction to changes in narrative attention is unrelated to fundamentals and is, therefore, detrimental to price informativeness. (iii) Narrative exposures proxy for this non-fundamental source of return variation and are negatively related to price informativeness in the cross-section. (iv) Narrative exposures proxy for a significant part of non-systematic return variance despite the fact that shocks to narrative attention explain only a modest fraction of return variance. (v) A shock to narratives' attention or bias boosts the trading volume in the stocks highly exposed to particular narratives.

Literature review. Our study is related to several developing and mature strands of literature, and we establish new and revealing connections among some research directions.

To quantify price informativeness empirically, we rely on the cross-sectional measure by Bai, Philippon, and Savov (2016), and we give a structural interpretation of this measure in our model. Recent studies have used this measure in various settings: Kacperczyk, Sundaresan, and Wang (2020) use it to analyze the effect of foreign institutional investments on price informativeness; Chen, Kelly, and Wu (2020) use it to measure information spillovers between buy-side and sell-side research, and Cao, Goyal, Ke, and Zhan (2022) use it to study the effect of options trading on stock price informativeness. Farboodi, Matray, Veldkamp, and Venkateswaran (2021) introduce a similar measure to quantify the effects of data abundance on the information content of prices. We contribute to this literature by relating price informativeness to return narrative exposures both theoretically and empirically.

[4] It is rather unimportant at which stage of the model the bias in the signal appears—directly in the narratives delivered by media outlets or in the processing of the narratives by investors. We use biased media as the primary channel for tractability and to better reflect the current state of the literature highlighting news media biases (e.g., Mullainathan and Shleifer, 2005; Baloria and Heese, 2018; Goldman, Martel, and Schneemeier, 2022).

This study also relates to the recent applications of news media text in economics and finance research. As in this study, Bybee, Kelly, Manela, and Xiu (2021) use LDA to quantify the structure of economic news and show that news predicts certain macro variables. Bybee, Kelly, and Su (2022) use LDA to extract latent risk factors from news text, and Hanley and Hoberg (2019) use the algorithm to study emerging risks in the financial sector. Other studies apply supervised or semi-supervised algorithms to infer certain economic quantities from news text. For instance, Baker, Bloom, and Davis (2016) develop an index of policy uncertainty, Manela and Moreira (2017) develop a news-based volatility index, Engle, Giglio, Kelly, Lee, and Stroebel (2020) construct a news-based climate risk measure, Liu and Matthies (2022) quantify investor concerns about economic growth, and Dim, Koerner, Wolski, and Zwart (2022) produce a news-implied sovereign default risk index. All of these studies focus on the role of the media as a valuable source of unstructured data relevant for tracking various economic quantities.

In contrast, we build on research highlighting news media biases (e.g., Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006; Reuter and Zitzewitz, 2006; Baloria and Heese, 2018; Goldman, Gupta, and Israelsen, 2021), as well as biases in investors' belief formation, such as over- and under-reaction (e.g., De Bondt and Thaler, 1985; Shleifer and Summers, 1990; Barberis, Shleifer, and Vishny, 1998; Frazzini, 2006; Bordalo, Gennaioli, Ma, and Shleifer, 2020), to document three main theoretically motivated results: (i) time-varying attention to specific narratives in the media affects firms heterogeneously; (ii) due to media and investor biases, firms that are disproportionately exposed to media narrative attention shocks have less informative stock prices; and (iii) exposure to high-frequency media attention shocks is a predominant driver of excess volatility in stock returns. Therefore, although the news media can yield useful signals, it distorts some firms' asset prices. We establish media attention to narratives as a theoretically sound and empirically important channel of disagreement in financial markets.

We also contribute to the literature on news media's effects on the stock market. Tetlock (2007) shows that media pessimism depresses the aggregate market return, consistent with models of noise and liquidity traders. Garcia (2013) shows that this destabilizing impact of the media is magnified in bad times. Calomiris and Mamaysky (2019) show that news predicts aggregate returns in a manner that suggests that news flow mainly captures non-priced risks. Tetlock, Saar-Tsechansky, and Macskassy (2008) show that sentiment in firm-specific news predicts returns. Some papers document stock return overreaction and/or underreaction to media coverage (e.g., Hillert, Jacobs, and Müller, 2014; Manela, 2014; Frank and Sanati, 2018). Our focus and approach markedly differ from these papers. While they primarily focus on the impact of the news media, mainly sentiment, on stock returns, we analyze the biases reflected in media narratives and establish theoretically and empirically the direct destabilizing impact of media narrative exposure on the information content of individual stock prices.

Our results provide important insights for the literature on demand-based asset pricing and the determinants of cross-sectional variance. Recent work (Koijen and Yogo, 2019, p.1488) estimates that changes in latent demand are the most important demand-side determinant of the cross-sectional variance of stock returns, explaining 81 percent of the cross-sectional variance. Gabaix and Koijen (2021) build on De Long, Shleifer, Summers, and Waldmann (1990)'s model that features noisy beliefs driving demand fluctuations. They identify changes in beliefs as one of the potential determinants of high-frequency flows. We find that stocks' exposure to narrative shocks is one such proxy for changes in beliefs that result in trading, in turn explaining over 85% of the total and idiosyncratic variances in the cross-section. Consistent with the proposed theoretical mechanism, we establish narrative exposure as the major characteristic explaining non-systematic variance in the cross-section of stocks, complementing the residual household income risk channel of Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016).

The rest of the study is organized as follows. Section 2 describes the data sources, construction of stock and firm characteristics, the extraction of narratives from news text, and the computation of media narrative exposures. It also contains the summary statistics of the main variables used for analysis in subsequent sections. Section 3 first analyzes how different stock return variance components (i.e., proxy for information channels affecting stock returns) relate to narrative exposure and, in turn, price informativeness, and then tests how price informativeness is affected by media narrative exposures directly. It also analyzes a link between narrative shocks and turnover. Section 4 develops a model that supports the empirical analysis. Section 5 concludes the paper.The online appendix contains a description of data processing procedures, as well as robustness tests and extensions.

2 Data and Variable Measurement

This section describes the main data sets and variables used in the study: Section 2.1 covers general stock variables, Section 2.2 defines the sources of news text, Section 2.3 describes the procedures for extracting narratives and measuring narrative exposure, and Section 2.4 provides summary statistics and a preliminary analysis. Our sample period spans from 1998-2021, because our news media data begins in 1998. Table A1 describes all of the variables used in this study.

2.1 Stock and Firm Characteristics

Our sample comprises US common stocks (share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ stock exchanges. We retrieve daily stock returns, prices, market capitalization, and volume from the daily data files of the Center for Research in Security Prices (CRSP). We obtain firm fundamentals from the Compustat North America Annual File. We exclude firms in the financial sector, firms with year-end market capitalization below $1 million, and filter out stock years with less than 20 observations and stock years in which a stock changed its primary exchange. We use daily factor returns from Kenneth French's Data Library with stock returns to compute factor exposures, idiosyncratic variance, and other characteristics.

We decompose stock return variances into components representing particular information channels using two approaches. First, each year we estimate from daily returns standard linear factor models (market model and four-, and five-factor models by Fama and French (1993), Carhart (1997), and Fama and French (2015)) to decompose excess returns into systematic and idiosyncratic components and compute their respective variances. Second, we decompose stock return variance into components stemming from market information, private information, public information, and noise using the vector autoregression framework of BNPW. We perform the decomposition separately for each stock yearly using daily returns. The details of the procedures for both approaches are provided in the Online Appendix OA.2.

2.2 News Media Text

Public information affecting agents' trading decisions flows primarily through the news media. For our purposes, one requires a news media outlet that is not only widely read by financial market participants but also has a relatively long history and is easily retrievable. We rely on the historical news archive of the WSJ for a large corpus of historical news text and use it to quantify the evolution of different media narratives and firms' exposure to those narratives.

We retrieve the WSJ's historical news archive through its website, spanning from 1998, the first year of availability, to 2021. We apply filters to remove sections of the Journal that are highly unlikely to be relevant to financial markets and that stand the chance of introducing unnecessary noise into our text corpora. These sections include Entertainment, Leisure & Arts, Sports, Lifestyle & Culture, and the like—in total, 37 categories. We further process the news article texts to reduce dimensionality and noise using the `SpaCy` text processing pipeline. We lemmatize words, convert text to lowercase, and exclude stopwords and entities such as persons, geopolitical areas, locations, and nationalities. We also exclude articles shorter than 20 words and end up with 348,649 news articles—averaging 1,206 articles per month—for further analysis.[5]

2.3 Extracting Media Narratives and Computing Narrative Exposures

Procedures for Extracting Media Narratives. Daily news text publications cover various issues that grab agents' attention and potentially shape various economic decisions, including stock trading. Such an information-rich environment has apparent benefits but poses significant challenges related to the extraction of the parsimonious set of narratives behind the news. However, as Shiller (2017) advocates, one can apply recent advances in textual analysis and natural language processing to extract the underlying topical narratives in news text.

We adopt the unsupervised machine learning Latent Dirichlet Allocation (LDA) algorithm of Blei, Ng, and Jordan (2003), which has been successfully used in settings similar to ours (e.g.,

[5] Bybee, Kelly, Manela, and Xiu (2021) and Bybee, Kelly, and Su (2022) also use the WSJ text corpus but have sample periods, starting from 1984, and a different number of news articles, roughly 764,000. The differences arise primarily because the authors obtained their text corpus directly from the Dow Jones Historical News Archive. In contrast, we only have access to digitally accessible online data.

Bybee, Kelly, Manela, and Xiu, 2021; Bybee, Kelly, and Su, 2022; Hanley and Hoberg, 2019). The implementation details are presented in Appendix OA.1.

We find a total of 33 narratives, which we manually label based on the (top-100) unigrams and bigrams with the largest rescaled term weights.[6] We aggregate the across-article narrative distribution daily to obtain the level of attention to each narrative on a given day as follows:

$$\theta_{l,\tau} = \frac{\frac{1}{M}\sum_{m=1}^{M}\theta_{m,l,\tau}}{D_\tau}, \tag{1}$$

where $\theta_{l,\tau}$ captures the level of attention to narrative l on day τ, $\theta_{m,l,\tau}$ denotes the level of attention to narrative l in article m on day τ, and $D_\tau = \sum_{l=1}^{L}\frac{\sum_{m=1}^{M}\theta_{m,l,\tau}}{M}$ is a normalization that ensures $\theta_{l,\tau}$ sums to one so that attention allocation each day is a probability distribution.

Quantifying Exposures to Media Narratives. We quantify firms' exposure to individual narratives by the weighted co-movement between stock returns and individual narrative attention shocks $\tilde{\theta}_{l,\tau}$, which are measured (similar to Bybee, Kelly, and Su, 2022) on day τ as the difference between day τ's attention level and the average attention level over the past five days ending on $\tau-1$, i.e., $\tilde{\theta}_{l,\tau} = \theta_{l,\tau} - \frac{1}{5}\sum_{i=1}^{6}\theta_{l,\tau-i}$. First, we estimate an augmented factor model for each firm n using daily stock returns in year t and narrative l's attention shocks:

$$r_{n,\tau} = \alpha + \beta_{n,t}^{\top}F_\tau + \beta_{n,t,l}^{narr}\tilde{\theta}_{l,\tau} + \varepsilon_{n,\tau}, \tag{2}$$

where $r_{n,\tau}$ is stock n's excess return, and F_τ is the vector of factor realizations (we use the four-factor Carhart (1997) model as the main specification) on day τ in year t. We define the aggregate *Narrative Exposure* as the average of individual narrative betas' absolute values weighted by the volatility of each narrative's attention each year t:[7]

$$Narrative\ Exposure_{n,t} = \frac{\sum_l |\beta_{n,t,l}^{narr}| \times \sigma_t(\theta_l)}{\sum_l \sigma_t(\theta_l)}. \tag{3}$$

[6] Table OB2 contains the 50 top terms for each narrative. We use the TF-IDF (Term Frequency–Inverse Dense Frequency) weighting, i.e., scale the narrative-term weights such that terms that occur very frequently in a given narrative but less so across all other narratives have high weights for that narrative.

[7] The attention levels are on average lowly correlated, and weighting by the covariance matrix of attention to narratives produces similar results.

Thus, $Narrative\ Exposure_{n,t}$ captures the average magnitude of stock n's return co-movement with attention shocks to all identified narratives.[8] Stocks with high exposure are potentially affected by trading decisions that move prices when any of the narratives witnesses strong attention shocks. Those trading decisions may be driven by the public information inherent in the attention shock or may be due to other information sources that are coincidentally manifested in the narrative attention shock.

To clarify, narrative exposure differs substantially from firm-specific news coverage. Although particular firms can receive high media coverage when certain narratives are actively discussed in the media, and as a result, smaller firms with typically low media coverage may experience higher narrative exposure when they are mentioned in the news compared to larger firms, we find that narrative exposure is barely correlated with firm-specific news coverage. Using firm-specific news coverage information from RavenPack, we compute the correlation between a firm's narrative exposure and the total number of mentions of the firm in the WSJ in a year within firm-size quintiles. The absolute correlations are close to zero (less than 0.09) and non-monotonic (slightly higher for mid-sized firms).

2.4 Summary Statistics and Preliminary Analysis

Table 1 shows the summary statistics for most of the variables used in the analysis, and Table OB1 in the Online Appendix shows correlations among variables of interest.

Figure 1 depicts the evolution of the aggregate $Narrative\ Exposure$ defined in (3), averaged each year across all firms in the sample, by size quintiles and major industry groups. In Panel A the market-wide narrative exposure demonstrates rich dynamics, clearly spiking a year or two before formal recessions in the economy (defined according to the National Bureau of Economic Research, NBER), and staying relatively low between these periods. Panel B reveals two striking and persistent patterns: (i) Exposure to media narratives decreases monotonically across size quintiles, which means that smaller firms' stock prices are generally more exposed to media narrative attention shocks. (ii) Exposure to media narratives spikes for firms across all size groups during major stock market downturns, but more so, again, for smaller firms. The first

[8] One can also define narrative exposure to each narrative l as $Narrative\ Exposure_{n,t,l} = |\beta^{narr}_{n,t,l}|$. We provide additional analysis using this definition for selected sub-groups of narratives in Online Appendix OC.

	Mean	Std	10%	25%	50%	75%	90%	Obs.
Panel A: Narrative exposure.								
$Narrative\ Exposure_{n,t}$	0.145	0.098	0.052	0.074	0.116	0.187	0.277	81,952
Panel B: Variance decomposition.								
$IdVar_{n,t} \times 10^3$	1.943	2.703	0.184	0.371	0.898	2.294	4.880	81,952
$SysVar_{n,t} \times 10^3$	0.218	0.334	-0.000	0.033	0.105	0.257	0.560	81,952
$MktInfo_{n,t} \times 10^3$	0.170	0.270	0.004	0.020	0.068	0.193	0.449	57,974
$PrivateInfo_{n,t} \times 10^3$	0.455	0.601	0.038	0.088	0.226	0.567	1.143	57,974
$PublicInfo_{n,t} \times 10^3$	0.737	1.038	0.070	0.137	0.333	0.880	1.863	57,974
$Noise_{n,t} \times 10^3$	0.852	1.498	0.042	0.104	0.293	0.855	2.180	57,974
Panel C: Factor model betas.								
$Market\ Beta_{n,t}$	0.858	0.528	0.113	0.503	0.889	1.211	1.538	81,952
$Size\ (SMB)\ Beta_{n,t}$	0.706	0.720	-0.206	0.162	0.645	1.179	1.743	81,952
$Value\ (HML)\ Beta_{n,t}$	0.138	0.804	-0.900	-0.326	0.134	0.615	1.116	81,952
$Mom\ (WML)\ Beta_{n,t}$	-0.105	0.578	-0.870	-0.421	-0.074	0.243	0.596	81,952
Panel D: Fundamentals and stock characteristics.								
$\ln(Assets)_{n,t}$	5.756	2.010	3.036	4.186	5.665	7.258	8.582	81,952
$\ln(Market\ Cap/Assets)_{n,t}$	0.036	0.942	-1.221	-0.611	0.038	0.700	1.312	81,952
$EBIT_{n,t}/Assets_{n,t}$	-0.019	0.218	-0.357	-0.055	0.053	0.107	0.168	81,952
$Debt_{n,t}/Assets_{n,t}$	0.209	0.199	0.000	0.011	0.170	0.347	0.516	81,952
$Cash_{n,t}/Assets_{n,t}$	0.227	0.244	0.010	0.034	0.126	0.349	0.654	81,952
$PP\&E_{n,t}/Assets_{n,t}$	0.237	0.219	0.027	0.064	0.158	0.349	0.623	81,952
$Sales_{n,t}/Assets_{n,t}$	0.971	0.686	0.157	0.447	0.846	1.361	2.001	81,952
$Capex_{n,t}/Assets_{n,t}$	0.047	0.046	0.006	0.015	0.032	0.063	0.113	81,952
$R\&D_{n,t}/Assets_{n,t}$	0.062	0.103	0.000	0.000	0.005	0.082	0.218	81,952
$Turnover_{n,t}$	8.622	9.138	1.445	3.007	6.126	11.184	18.891	81,952
$Illiquidity_{n,t}$	0.319	0.861	0.001	0.002	0.013	0.112	0.951	81,883
$MAX_{n,t}$	0.537	0.278	0.134	0.312	0.553	0.774	0.908	81,952
Panel E: Institutional variables.								
$DOB_{n,t}$	0.007	0.014	0.001	0.001	0.003	0.007	0.018	23,689
$Inst.\ Ownership_{n,t},\ \%$	0.506	0.320	0.050	0.202	0.536	0.791	0.920	66,887

Table 1: Summary Statistics.
The table shows the summary statistics for selected variables computed from the firm-year panel data. Each year, all continuous variables are winsorized at 5% and 95% levels.

pattern serves as an initial piece of evidence consistent with our theoretical framework. We expect media biases or decisions of agents with biased interpretations of news media coverage to have a more profound impact, for instance, through trading, on the stock prices of smaller firms, leading to the observed higher exposure to narrative attention shocks for such firms. This is because smaller firms are more likely to be traded by investor groups with a higher tendency to exhibit behavioral biases (e.g., Barber and Odean, 2000), and, at the same time, it is harder for rational agents to exploit such biases due to limits to arbitrage.

Similarly, the spike in narrative exposure across firms in bad times is consistent with existing evidence that news media impacts aggregate stock market prices, particularly in recessions

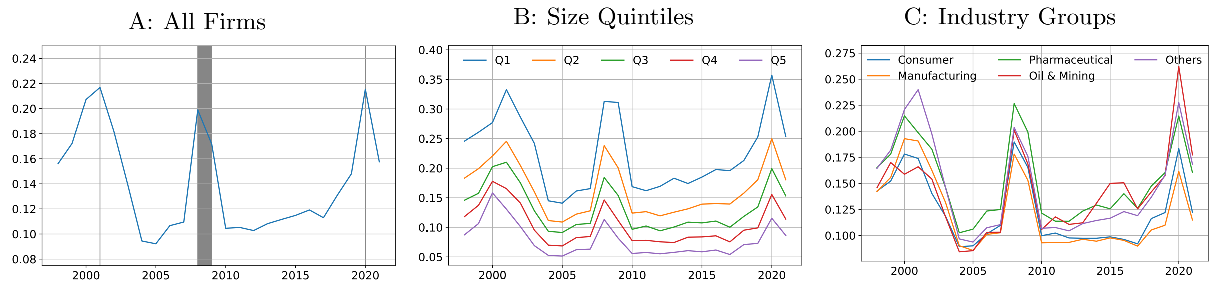

Figure 1: Media Narrative Exposure. The figure shows the evolution of the aggregate $Narrative\ Exposure_{n,t}$, as defined in equation (3), averaged across all firms (Panel A), size quintiles (Panel B) and industry groups (Panel C). In Panel C, the Fama-French 17 industries are collapsed into five major groups to facilitate exposition. The *Consumer* group comprises the Food, Clothing, and Consumer Durables industries; the *Manufacturing* group comprises the Construction, Steel, Fabricated Products, Machinery, and Utilities industries; the *Pharmaceutical* group comprises the Chemicals and Consumer Drugs industries; the *Oil & Mining* group comprises the Mines, Oil, and Steel industries; and the *Others* group comprises the remaining industries. Panel A contains shaded areas indicating NBER recession periods.

(Garcia, 2013). Here, we further document that in the cross-section, the news media's tendency to distort prices in bad times is likely more pronounced for smaller firms, since such firms are more exposed to media narratives, and their exposures spike even more disproportionately during market downturns.

Panel B of Figure 1 reveals that firms' exposure to media narratives is not driven by some specific industry group. For example, in the early sample period, the Oil & Mining industry group had one of the lowest average exposures but had one of the largest exposures by the end of the sample. The figure further indicates that media narrative exposure exhibits similar time-series trends across industries—again, commonly surging during market downturns. This evidence illustrates that the extracted media narrative exposures are not merely artifacts of estimation error or random fluctuations. Even though they are estimated individually for each firm, we observe strong commonality over time across groups of stocks.

3 Narratives, Information Channels and Price Informativeness

This section establishes the empirical link between narrative exposure, information channels in stock returns and informativeness of prices. Section 3.1 shows that narrative exposure is closely linked to idiosyncratic and especially public information-related part of non-systematic variance. Section 3.2 establishes how the levels of non-systematic variances interact with price informativeness, and then Section 3.3 examines how exposure to media narratives affects price informativeness regarding future firm fundamentals directly. Finally, Section 3.4 analyzes a link between narrative exposure and trading volume.

3.1 Information Channels in Stock Returns and Narrative Exposure

For analyzing the channels through which information is getting into the stocks' prices and driving their returns, we use two sets of proxies, clearly separating variances into systematic and non-systematic components.[9] The first is a combination of systematic ($SysVar$) and idiosyncratic variances ($IdVar$) estimated from standard factor models. Systematic variance captures market-wide information that jointly affects all individual firms' stock prices and is not particularly informative regarding an individual firm's future cash flow. Conversely, idiosyncratic variance stems from at least three sources: (i) firm-specific information not reflected in the aggregate market dynamics, (ii) agents' heterogeneous interpretation of how public information deferentially affects firms, and (iii) noise trading unrelated to either public or firm-specific information. The relationship between the level of idiosyncratic variances and the corresponding asset prices' informativeness will likely depend on which of these sources of idiosyncratic price variation is dominant for specific stocks.

Our second set of information channels targets a different and more granular decomposition of stock return variation, allowing for a finer separation of the components of idiosyncratic variance. Precisely, we use the framework of Brogaard, Nguyen, Putnins, and Wu (2022) (BNPW henceforth) to decompose total stock return variance into components stemming from market-wide ($MktInfo$), private ($PrivateInfo$) or public ($PublicInfo$) firm-specific information, and noise ($Noise$). $MktInfo$ is similar to $SysVar$ from a factor model but is identified using vector autoregression as the response of stock returns to market factor shocks only. Private and public firm-specific information are respectively identified as a permanent stock return response to trading volume and own-return shocks after controlling for market return shocks. Noise absorbs the residual variance.[10]

[9]Note that partitions of return variance into components provide us a view of the intensity of information channels driving stock returns.

[10]BNPW note that "in reality, the distinction between public and private information can sometimes be blurred," so we refrain from drawing strong conclusions based on this distinction.

While we use the levels of the variance components for our analysis, we examine their proportions to determine whether they are comparable to those of BNPW. In factor-based models, the share of the average systematic variance in total variance ranges from 8.5% for the one-factor model to 11.5% for the five-factor model. Clearly, the residual idiosyncratic variance share is, on average, very large. The numbers for the BNPW decomposition are roughly comparable to the original study, even though we have a shorter (and later) sample period (1998 to 2021 compared to 1960 to 2015 in BNPW). We find that market-wide information accounts for 7.4% of the return variance, private information accounts for 20.2% of the variance, public information accounts for 32.3%, and the remaining 40.1% is noise. The respective numbers from an earlier sample in BNPW are 8%, 24%, 37%, and 31%, respectively. Consistent with BNPW, we find a decreasing trend in noise variance for most of the sample period and an increasing trend for firm-specific information. However, a sharp increase in the noise component and an equivalent drop in the firm-specific (mostly public) variance in 2020-2021 lead to a slight discrepancy in proportions.

	$SysVar_{n,t}$	$IdVar_{n,t}$	$MktInfo_{n,t}$	$PrivateInfo_{n,t}$	$PublicInfo_{n,t}$	$Noise_{n,t}$
$SysVar_{n,t}$	1.000	0.043	0.550	0.134	0.096	0.002
$IdVar_{n,t}$	0.043	1.000	0.342	0.783	0.890	0.841
$MktInfo_{n,t}$	0.550	0.342	1.000	0.362	0.407	0.184
$PrivateInfo_{n,t}$	0.134	0.783	0.362	1.000	0.722	0.502
$PublicInfo_{n,t}$	0.096	0.890	0.407	0.722	1.000	0.643
$Noise_{n,t}$	0.002	0.841	0.184	0.502	0.643	1.000

Table 2: Correlation of Information Channels.
The table provides unconditional correlations of information channel proxies for individual stocks: systematic and idiosyncratic variances based on the four-factor model and BNPW variance decomposition. The sample period is from 1998 to 2021, with annual frequency. All proxies are computed, winsorized at 5% and 95%, and are then standardized to unit variance on an annual basis.

Table 2 shows the correlation of the information channel proxies. The systematic and non-systematic information sources do not overlap much across the two methodologies, but the factor-based systematic variance is somewhat correlated (0.55) with systematic variance from BNPW. On the other hand, the factor-based idiosyncratic variance is highly correlated with all three non-systematic variance components from BNPW (correlations of 0.8-0.9). We see that all of the non-systematic variance components are jointly driven by some common factors or characteristics, and the intensities of the information channels they reflect are strongly connected.

To test for a link between absolute narrative exposure and the non-systematic variance measures in the cross-section of stocks, we use a two-stage procedure, in which we annually regress each variance component on stocks' $Narrative\ Exposure$ while controlling for a host of stock characteristics, and then average the time-series coefficients. We control for a large set of traditional characteristics so as to isolate the relevance of media narrative exposure from other variables that are potentially relevant to cross-sectional differences in the variance components.

The results are provided in Table 3. The full specification in Panel A includes $Narrative\ Exposure$, four-factor betas, fundamental variables, stock characteristics, and sector fixed effects. The reduced specification in Panel B contains only the $Narrative\ Exposure$. All continuous variables on both sides are winsorized annually at 5% and 95%, and are then standardized to have a cross-sectional variance of one. The results are truly striking. Comparing the estimates in the specifications in Panels A and B, we see that in terms of economic magnitude, media narrative exposure is the single most important driver of non-systematic variance components in stock returns. More so, media narrative exposure alone explains a whooping 86% of the variation in idiosyncratic variance and 59%-71% of the variation in variances due to public and private information and noise components.

	$Var_{n,t}$	$SysVar_{n,t}$	$IdVar_{n,t}$	$MktInfo_{n,t}$	$PrivateInfo_{n,t}$	$PublicInfo_{n,t}$	$Noise_{n,t}$
Panel A: Full Specification.							
$Narrative\ Exposure_{n,t}$	0.776	-0.042	0.795	0.208	0.632	0.629	0.646
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	87.75	77.96	87.96	48.81	64.19	74.78	66.05
Obs.	2,413	2,413	2,413	2,413	2,413	2,413	2,413
Factor betas	FF4	FF4	FF4	FF4	FF4	FF4	FF4
Fundamentals	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Reduced Specification.							
$Narrative\ Exposure_{n,t}$	0.923	0.052	0.928	0.359	0.764	0.845	0.777
	(0.001)	(0.208)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	85.17	2.28	86.09	13.97	58.51	71.37	60.49
Obs.	2,413	2,413	2,413	2,413	2,413	2,413	2,413
Controls/ FE	No	No	No	No	No	No	No

Table 3: Information Channels and Firm Characteristics.
The table shows the cross-sectional link between the intensity of information channels driving individual stock returns and firm characteristics. Information channels are systematic and idiosyncratic variances based on the four-factor model and BNPW variance decomposition. The coefficients are based on the two-stage regression. Panel A shows results with all regressors and sector dummies control, and Panel B shows a reduced specification without controls. The sample period is from 1998 to 2021, with annual frequency. All continuous variables are winsorized at 5% and 95%, and are then standardized to unit variance in the cross-section on an annual basis. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and the number of observations (Obs.) are average numbers from the cross-sectional stage.

For example, a one-standard-deviation (STD) increase in the narrative exposure is linked to a $0.93\times STD$ increase in the idiosyncratic variance $IdVar$ in Panel B, and to a $0.80\times STD$ increase after controlling for all other characteristics in Panel A. The reduced specification's R^2 of 86% increases by less than 2% in the full specification. The $PublicInfo$ column shows a similar pattern: a $1\times STD$ increase in the average absolute narrative beta is linked to $0.85\times STD$ and a $0.63\times STD$ increase in the variance due to public information for the reduced and full specifications, respectively. The R^2's are 75% and 71% for the full and reduced specifications, respectively. $Noise$ and $PrivateInfo$ are slightly less strongly related to narrative exposure. $MktInfo$ is statistically linked to narrative exposure, but the economic magnitude is relatively negligible. The factor-based systematic variance is not positively related to narrative exposure.

Thus, stocks highly exposed to media narratives also have high levels of idiosyncratic variance linked to (and potentially explained by) high variance due to trading on public information and noise produced by the news media. In the next section, we directly test whether there is a statistical link between idiosyncratic variance, variances due to public information and noise on the one side, and price informativeness on the other.

3.2 Information Channels and Price Informativeness

To determine how various information channels empirically relate to the information content of stock prices with respect to future fundamentals, we adopt a stock-level measure of price informativeness based on Bai, Philippon, and Savov (2016), defined as the predicted variation of cash flows by current market prices. More precisely, we test whether higher intensity of a particular information flow makes current stock prices less informative about future firm fundamentals. Our main model is specified as the Fama and MacBeth (1973) regression of future earnings h years from today relative to current assets, $E_{n,t+h}/A_{n,t}$, on current earnings, market value relative to assets, $\ln(M_{n,t}/A_{n,t})$, the interaction of market value and particular information channels, and controls:

$$\frac{E_{n,t+h}}{A_{n,t}} = a + b_{0,h}\frac{E_{n,t}}{A_{n,t}} + \left[b_{1,h} + b_{proxy,h}^{\top} proxy_{n,t}\right] \ln\frac{M_{n,t}}{A_{n,t}} + b_x^{\top} X_{n,t} + \varepsilon_{n,t+h}, \tag{4}$$

where h is one or three years, $proxy_{n,t}$ denotes a vector with information channel proxies of firm n, and the vector of controls, $X_{n,t}$, includes the information channel proxy used in interaction term, four-factor model betas, fundamental variables, namely, $\ln(Assets)$, $Debt/Assets$, $Cash/Assets$, $Ppent/Assets$, $Capex/Assets$, $Sales/Assets$, $R\&D/Assets$, and economic sector dummies (eight one-digit SIC codes after excluding the financial sector). All continuous variables are winsorized at 5% and 95% for each year in the sample period. The market value variable $\ln(M/A)$ is standardized to unit variance each year in the cross-section so that the coefficient, $b_{1,h}$, directly provides the proxy for price informativeness following Bai, Philippon, and Savov (2016). The coefficient $b_{2,h}$, therefore, reveals how price informativeness interacts with a particular information channel. In terms of information channel proxies, we use the two sets of variance decomposition, factor-based variances, and variances from VAR estimation in BNPW.

	One-year horizon				*Three-year horizon*			
	MM	FF4	FF5	BNPW	MM	FF4	FF5	BNPW
$\ln(M/A)_{n,t}$	0.019	0.019	0.020	0.021	0.035	0.038	0.038	0.040
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times SysVar_{n,t}$	-0.000	-0.001	-0.002	–	0.000	-0.000	-0.000	–
	(0.521)	(0.070)	(0.028)		(0.992)	(0.977)	(0.892)	
$\ln(M/A)_{n,t} \times IdVar_{n,t}$	-0.014	-0.014	-0.014	–	-0.020	-0.024	-0.024	–
	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	
$\ln(M/A)_{n,t} \times MktInfo_{n,t}$	–	–	–	-0.003	–	–	–	0.001
				(0.001)				(0.797)
$\ln(M/A)_{n,t} \times PrivateInfo_{n,t}$	–	–	–	-0.003	–	–	–	0.008
				(0.001)				(0.556)
$\ln(M/A)_{n,t} \times PublicInfo_{n,t}$	–	–	–	-0.009	–	–	–	-0.014
				(0.001)				(0.001)
$\ln(M/A)_{n,t} \times Noise_{n,t}$	–	–	–	-0.003	–	–	–	-0.007
				(0.001)				(0.076)
R^2 (%)	79.68	79.69	79.69	80.41	60.75	60.86	60.86	62.56
Obs.	3,151	3,151	3,151	2,223	2,470	2,470	2,470	1,736
Factor betas	FF4	FF4	FF4	FF4	FF4	FF4	FF4	FF4
Fundamentals	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: Information Channels and Price Informativeness.
The table shows aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its interaction with information channel proxies. The model is estimated as the Fama-MacBeth regression (4) for one- and three-year horizons. The first three columns of each horizon use factor models (market, four- and five-factor models) for variance decomposition into systematic ($SysVar_{n,t}$) and idiosyncratic ($IdVar_{n,t}$) components, and column $BNPW$ uses the decomposition of Brogaard, Nguyen, Putnins, and Wu (2022). Controls include four-factor model betas, fundamental and stock characteristics, and sector dummies. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and standardized to unit variance. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and number of observations (Obs.) are average numbers from the cross-sectional stage.

The results in Table 4 demonstrate that while stock prices are, on average, informative about future fundamentals for horizons of one and three years, price informativeness significantly decreases for stocks with high levels of idiosyncratic variance. The effect is economically large, and for the one-year horizon, $1 \times STD$ difference in $IdVar$ decreases the price informativeness by 70% (adjustment of -0.014 applied to the base level of ≈ 0.020). The levels of systematic variance in most cases do not significantly affect price informativeness (except for the five-factor model with an economically low magnitude of -0.002 for an interaction term). In all cases, the interaction term for the $SysVar$ is approximately an order of magnitude smaller than for the $IdVar$. The results for all factor models in the table are similar.[11]

With a more granular variance decomposition (in column BNPW), we observe for both horizons the largest and most significant decrease in price informativeness for stocks with high $PublicInfo$ variance. Keeping market value constant, a $1 \times STD$ change in $PublicInfo$ decreases price informativeness about future one-year fundamentals by around 50% (i.e., by 0.009 compared to the base level of 0.021). $Noise$ also significantly drives price informativeness in the same direction, with the economic magnitude roughly 3 times smaller. $PrivateInfo$ and $MktInfo$ are also statistically significant, but economically, their contribution is small as well. For the three-year horizon, $PublicInfo$ is the only information channel significantly interacting with price informativeness, but at 1% significance level. Interaction with the noise component is economically sizeable but only borderline significant (with p-value of 0.076). Overall, price informativeness is negatively associated with non-systematic variance, and the effect is primarily driven by public information.

3.3 Narratives and Price Informativeness

To determine whether exposure to media narratives directly affects the information content of stock prices, we continue measuring the predicted variation of cash flows by current market prices. More precisely, we test whether high exposure to media narratives results in lower stock price informativeness about future firm cash flows by regressing future earnings h years from today relative to current assets, $E_{n,t+h}/A_{n,t}$, on current earnings, market value relative to assets,

[11] In the subsequent analysis, we select the Carhart (1997) four-factor model as our benchmark and check the sensitivity to other factor models in terms of robustness.

$\ln(M_{n,t}/A_{n,t})$, the interaction of market value and particular narrative exposure, and controls:

$$\frac{E_{n,t+h}}{A_{n,t}} = a + b_{0,h}\frac{E_{n,t}}{A_{n,t}} + [b_{1,h} + b_{2,h} \times \mathit{Narrative\ Exposure}_{n,t}] \ln \frac{M_{n,t}}{A_{n,t}} + b_{x,h}^{\top} X_{n,t} + \varepsilon_{n,t+h}, \quad (5)$$

where h is one or three years, $\mathit{Narrative\ Exposure}_{n,t}$ is defined in equation (3), and the vector of controls is as in the previous section.[12]

Table 5 shows that price informativeness significantly decreases for stocks with high narrative exposure for both the one- and three-year horizons. This result delivers a profound message: firms whose stock prices co-vary substantially with media narratives, in general, tend to absorb irrelevant information that renders prices uninformative. The loss of price informativeness arises from the inherent media bias that, when traded upon, tends to distort affected firms' stock prices.

	One-year horizon				*Three-year horizon*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln(M/A)_{n,t}$	0.022	0.032	0.032	0.032	0.046	0.059	0.058	0.060
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times \mathit{Narrative\ Exposure}_{n,t}$	-0.016	-0.015	-0.015	-0.009	-0.028	-0.025	-0.024	-0.016
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\mathit{Illiquidity}_{n,t}$	–	–	0.001	-0.001	–	–	0.001	-0.001
			(0.462)	(0.001)			(0.711)	(0.179)
$\mathit{MAX}_{n,t}$	–	–	-0.001	0.009	–	–	0.009	0.012
			(0.784)	(0.001)			(0.178)	(0.001)
$\mathit{DOB}_{n,t}$	–	–	–	-0.010	–	–	–	-0.007
				(0.001)				(0.001)
$\mathit{Inst.\ Ownership}_{n,t}$, %	–	–	–	0.003	–	–	–	0.005
				(0.001)				(0.005)
R^2 (%)	77.94	79.40	79.46	77.54	57.04	60.31	60.50	55.28
Obs.	3,151	3,151	3,151	946	2,470	2,470	2,470	859
Factor betas	–	FF4	FF4	FF4	–	FF4	FF4	FF4
Fundamentals	–	Yes	Yes	Yes	–	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High Average Exposure	-0.006	-0.006	-0.006	-0.003	-0.011	-0.004	-0.003	-0.001
	(0.013)	(0.015)	(0.014)	(0.038)	(0.306)	(0.443)	(0.472)	(0.853)

Table 5: Price Informativeness and Narrative Exposure.
The table shows aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its interaction with $\mathit{Narrative\ Exposure}_{n,t}$ defined in (3). The model is estimated as the two-stage regression (5) for one- and three-year horizons. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and market value $\ln(M/A)$ is standardized to unit standard deviation. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. $R^2(\%)$ and the number of observations (Obs.) are average numbers from the cross-sectional stage.

[12] All continuous variables are winsorized at 5% and 95% for each year in the sample period. The market value variable $\ln(M/A)$ is standardized to unit variance each year in the cross-section so that the coefficient, $b_{1,h}$, directly provides the proxy for price informativeness. The coefficient $b_{2,h}$ reveals how price informativeness interacts with the narrative exposure.

At the end of Table 5, we estimate the marginal change in the incremental price informativeness, i.e., the interaction term, conditional on periods of high average narrative exposure. For this, we regress the time-series of the interaction term coefficient $b_{2,h}$ from the cross-sectional stage of the Fama-MacBeth procedure on a constant and a dummy variable that equals one for the years of high average narrative exposure, defined as periods when the cross-sectional mean of $Narrative\ Exposure_{n,t}$ is above its sample mean, and zero otherwise. We report the coefficient on the dummy variable along with its $p-$value. For both horizons, high levels of the average narrative exposure exacerbate the loss of price informativeness, though the results are significant only for the one-year regression.

Thus, a high level of narrative exposure is *associated* with lower stock price information content, and the effect is stronger during periods of elevated average narrative exposure in the market. While this result is insightful, there is a likely endogeneity concern that may arise from the level of narrative exposure being correlated with certain firm characteristics that are equally related to stock price informativeness. If this is the case, then the lower price informativeness of high narrative exposure firms could arise from other reasons unrelated to the level of narrative exposure. Although the stability of our estimated coefficients across different horizons and specifications with different sets of control variables suggests that this is unlikely the case, it does not fully address the endogeneity concern.

Ideally, the endogeneity concern can be resolved using a natural experiment that generates exogenous variation in narrative exposure without directly affecting price informativeness. In practice, however, true natural experiments are uncommon, and researchers resort to quasi-natural experiments that plausibly generate the desired exogenous variation. We follow a similar approach to get closer to causality by comparing firms that witnessed a sudden large increase in narrative exposure to other firms similar across observable characteristics but did not experience the same large increase in narrative exposure. Implicitly, we assume that among similar firms, the sudden and large increase in narrative exposure experienced by some firms is due to reasons other than their price informativeness. Such an increase in narrative exposure could arise due to the news media's choice to reallocate attention across topical narratives. This impacts what investors glean from the news and trade on, ultimately affecting firms' exposure to media narratives heterogeneously.

We proceed by first examining the persistence in narrative exposure, as it plays a role in what one could consider a large change in yearly exposure levels. Table 6 shows the average migration matrix across narrative exposure quintiles from a given year $t-1$ to t. The value in the top left corner indicates that a stock ranked in the bottom narrative exposure quintile in year $t-1$ is 66.2% likely to remain in that quintile in t. On the other hand, the value in the top right corner indicates that the same stock has only a 0.8% probability of transitioning to the top quintile in t. Hence, there is reasonable persistence in narrative exposure across adjacent years. We, therefore, define a large increase in narrative exposure as an increase in a firm's narrative exposure percentile rank by at least 25 percentage points between year $t-1$ and t.[13] We use firms that experienced such an increase in narrative exposure ranking as the "treated firms", yielding 3,664 unique treated firms—out of which 62% are treated once—over the full sample. On average, there are 248 treated firms per year.

New Old	1	2	3	4	5
1	0.662	0.241	0.069	0.020	0.008
2	0.248	0.383	0.238	0.100	0.031
3	0.072	0.253	0.348	0.237	0.090
4	0.015	0.103	0.257	0.367	0.258
5	0.003	0.020	0.089	0.275	0.613

Table 6: Narrative Exposure Migration Matrix.
The table shows the average proportion of firms migrating from an old to a new $Narrative\ Exposure_{n,t}$ quintile from one year to the next. $Narrative\ Exposure_{n,t}$ is computed following equation (3).

Next, for each treated firm, we identify up to five control firms that are similar across observable characteristics using the propensity score matching algorithm. Our matching is based on the following characteristics observed one year before treatment: $Narrative\ Exposure$, $\ln(Market\ Cap.)$, $\ln(Market\ Cap./Assets)$, $\ln(BTM)$, $EBIT/Asset$, $Capex/Assets$, $R\&D/Assets$, $Market\ Beta$, and $Illiquidity$. We then estimate the following panel regression using the sample of treated and matched control firms:

$$\frac{E_{n,t+h}}{A_{n,t}} = a + b_{0,h}\frac{E_{n,t}}{A_{n,t}} + [b_{1,h} + b_{2,h} \times Treated_{n,t}] \ln \frac{M_{n,t}}{A_{n,t}} + b_{x,h}^{\top} X_{n,t} + \delta_t + \varepsilon_{n,t+h}, \tag{6}$$

[13] We use 25 percentage points threshold due to the moderate-high persistence level in narrative exposure and to ensure a sizeable number of treated firms necessary for the test power. In the Online Appendix OC we vary the threshold to 20 and 30 percentage points, finding similar results.

where h is one or three years, $Treated_{n,t}$ is a dummy variable that equals 1 for firms that experienced an increase in narrative exposure percentile rank of at least 25 percentage points from year $t-1$ to t, and δ_t captures year fixed effects. The vector of control variables, $X_{n,t}$, allows us to account for any differences in observable characteristics across the two groups of firms. These control variables are $Treated_{n,t}$, fundamental variables ($\ln(Assets)$, $Debt/Assets$, $Cash/Assets$, $PP\&E/Assets$, $Capex/Assets$, $Sales/Assets$, $R\&D/Assets$), stock characteristics (four-factor model betas, $Turnover$, $Illiquity$, MAX), and economic sector dummies (eight one-digit SIC codes after excluding the financial sector).

Our coefficient of interest is $b_{2,h}$, which captures the change in price informativeness for the treated group relative to the control group. Table 6 reports the estimates of $b_{1,h}$ and $b_{2,h}$, clearly showing that for the control group, stock prices are significantly informative about future firm cash flow. Conversely, the interaction term is significantly negative, indicating that relative to their similar peers, price informativeness significantly declines for firms that experience a large increase in narrative exposure. The effect is economically sizeable and statistically significant across specifications for the one-year and three-year horizons. For example, the coefficients reported in columns (4) and (8) indicate that price informativeness about cash flows realized over the next one and three years decreases by roughly 42% for the treated relative to the control firms.

	One-year horizon				*Three-year horizon*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln(M/A)_{n,t}$	0.0112	0.0109	0.0109	0.0110	0.0192	0.0220	0.0222	0.0222
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(M/A)_{n,t} \times Treated$	-0.0045	-0.0048	-0.0046	-0.0046	-0.0103	-0.0097	-0.0095	-0.0094
	(0.046)	(0.022)	(0.032)	(0.035)	(0.000)	(0.000)	(0.000)	(0.000)
R^2 (%)	70.79	71.22	71.37	71.96	45.46	46.92	47.15	48.58
Obs.	34,350	34,350	34,350	34,350	25,722	25,722	25,722	25,722
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	No	Yes	No	Yes	No	Yes	No
Year FE	No	Yes	Yes	No	No	Yes	Yes	No
Sector x Year FE	No	No	No	Yes	No	No	No	Yes

Table 7: Narrative Exposure Changes and Price Informativeness.
The table shows the aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and the change in price informativeness for firms (coefficient for interaction with $Treated$) that experienced a large annual change (at least 25 percentage points increase in percentile rank) in the narrative exposure relative to comparable firms. The panel fixed effect regression (6) is estimated for the one- and three-year horizons (columns 1-4 and 5-8, respectively). Controls include four-factor betas, fundamental and stock characteristics, and various fixed effects described under Eq. (6). The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and market value $\ln(M/A)$ is standardized to unit standard deviation. $p-$values for coefficients in parentheses use standard errors clustered at year-firm level.

Overall, although the preceding analysis does not rule out all potential endogeneity issues, it does lend more confidence to a causal interpretation of the results. The evidence is in line with the baseline analysis, suggesting that high stock price exposure to media narrative attention shocks weakens price informativeness about future fundamentals. This constrains the information agents can glean from asset prices with potential adverse implications for the economy-wide allocative efficiency (see, Bond, Edmans, and Goldstein, 2012).

3.4 Narrative Exposure and Trading Activity

Following our earlier intuition, shocks to media narrative attention change the information available to agents, leading to updates in stock return expectations and the subsequent adjustments in portfolio holdings. Thus, stocks affected more strongly by narrative attention shocks should experience higher turnover. We test this claim by relating average turnover to narrative exposure, controlling for a number of other variables that potentially affect market activity. We continue working on the annual frequency with the two-stage framework as in the previous sections. Turnover is computed as the yearly average of the ratio of trading volume (number of shares traded) to the total number of shares outstanding.

The results in Table 8 confirm that an increasing narrative exposure is associated with an elevated trading volume—the $Narrative\ Exposure_{n,t}$ coefficient is always positive and significant. In the first column for the regression without continuous controls, the average first-stage R^2 is 11.1%, and $1 \times STD$ higher narrative exposure corresponds to $0.2 \times STD$ higher relative turnover. Adding various controls for the same sample (up to column 4 in the Table) boosts the explanatory power of the cross-sectional stage, and also increases the slope of the average narrative shock, which hints at potential interaction between regressors. For a smaller sample of stocks (in column 5), $1 \times STD$ higher narrative exposure corresponds to $0.4 \times STD$ higher relative turnover after controlling for all other characteristics.

Thus, we find that high narrative exposure is linked to reduced stock price informativeness, especially when the average stock exposure to narratives is high; moreover, the prices of stocks with increasing narrative exposure become less aligned with firms' future fundamentals compared to the prices of their peers not experiencing an increase in the narrative exposure. We also

	$Turnover_{n,t}$				
	(1)	(2)	(3)	(4)	(5)
$Narrative\ Exposure_{n,t}$	0.201	0.198	0.376	0.334	0.410
	(0.010)	(0.006)	(0.001)	(0.001)	(0.001)
$Market\ Beta_{n,t}$	–	0.378	0.234	0.174	0.079
		(0.001)	(0.001)	(0.001)	(0.026)
$Size\ (SMB)\ Beta_{n,t}$	–	0.039	0.074	0.025	-0.073
		(0.203)	(0.001)	(0.256)	(0.002)
$Value\ (HML)\ Beta_{n,t}$	–	-0.156	-0.137	-0.127	-0.085
		(0.004)	(0.001)	(0.001)	(0.014)
$Mom.\ (WML)\ Beta_{n,t}$	–	0.005	0.026	0.021	0.033
		(0.898)	(0.385)	(0.443)	(0.320)
$Illiquidity_{n,t}$	–	–	–	-0.202	-1.001
				(0.001)	(0.001)
$MAX_{n,t}$	–	–	–	0.108	0.354
				(0.001)	(0.001)
$DOB_{n,t}$	–	–	–	–	0.048
					(0.001)
$Inst.\ Ownership_{n,t},\ \%$	–	–	–	–	0.489
					(0.001)
R^2 (%)	11.10	30.32	39.31	42.24	48.28
Obs.	3,412	3,412	3,412	3,412	980
Controls:					
Fundamentals	No	No	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes

Table 8: Narrative Exposure and Trading Volume.
The table shows the cross-sectional link between the turnover ($Turnover$) defined as the average of the ratio of trading volume relative to shares outstanding, and $Narrative\ Exposure$, controlling for firm and stock characteristics, including sector dummies (SIC1-code) and in the last two specifications also the dispersion of beliefs (DOB) and institutional ownership. The coefficients are based on the two-stage Fama–MacBeth regression. The sample period is from 1998 to 2021, with annual frequency. For the specification with the institutional ownership, the sample period is from 1999 to 2018. All variables except for industry dummies are winsorized at 5% and 95%, and are then standardized to unit variance in the cross-section on an annual basis. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. $R^2(\%)$ and the number of observations (Obs.) are average numbers from the cross-sectional stage.

observed a strong correlation between high narrative exposure and elevated non-systematic variance and turnover rates across stocks. These findings hold even after considering various relevant characteristics.

Robustness and extensions. We test the sensitivity of our results to various modifications in procedures in the Online Appendix OC. In summary, price informativeness results in Table 4 remain strong if we remove the years containing NBER recessions (2001, 2008-2009, and 2020) from the sample period. The analysis of the changes in narrative exposure and the price informativeness in Table 7 is robust to varying the narrative exposure change threshold for defining the treated firms from 25 to 20 or 30 percentage points. Using the absolute narrative betas for selected narrative groups as the narrative exposure (instead of the weighted average absolute beta in the main analysis) shows that elevated exposure to individual narratives equally harms price informativeness. The result holds even for narratives that a priori can hardly be linked to the fundamentals of all firms in the market (e.g., schooling or entertainment).

4 A Model of Media Narratives and Price Informativeness

This section provides a stylized model to rationalize our empirical findings on narrative exposure and price informativeness, and provide a framework for their interpretation. Section 4.1 provides an informal overview of the model's assumptions and predictions, emphasizing the links to the empirical findings from earlier sections. Section 4.2 and Section 4.3 respectively set up and solve the model, formally deriving its predictions.

4.1 Model Overview

The model rests on two assumptions. The initial one, supported by extensive theoretical and empirical literature, asserts that the media exhibits bias (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006; Reuter and Zitzewitz, 2006; Baron, 2006; Goldman, Gupta, and Israelsen, 2021), and this media bias influences consumers' beliefs and actions, particularly among those unaware of the bias (Reuter and Zitzewitz, 2006; Baloria and Heese, 2018; DellaVigna and Kaplan, 2007; Gurun and Butler, 2012). The second one is that media communicates information through narratives that are informative about asset cash flows, with media narrative attention varying over time. This assumption directly links the information structure in the model to the empirical framework for media narratives in Section 2.3.

We integrate these assumptions into a stylized dynamic trading model with rational and unsophisticated investors who are unaware of the media bias, and we derive the following results. First, our model provides a mechanism for the covariance between narrative attention and stock returns (Proposition 1-(iii)) that supports our empirical definition of narrative exposure in Eq. (3). Essentially, as attention to a narrative increases, the associated bias influences unsophisticated investors' beliefs more heavily. As asset prices reflect these beliefs, stock returns move according to the narrative bias. This mechanism requires all our assumptions: without media bias, with investors fully aware of the bias, or if media narrative attention remains constant, all stocks' narrative exposures become nil.

Second, the stock price response to fluctuations in narrative attention, driven by bias, is independent of fundamentals, thereby diluting price informativeness (Proposition 2-(i)). This prediction aligns with the results in Section 3.2, showing that idiosyncratic volatility is inversely related to price informativeness.

Third, narrative exposures act as proxies for this non-fundamental source of return variation and are inversely related to price informativeness in the cross-section (Proposition 2-(ii)). Hence, the model provides a conceptual framework for interpreting the evidence in Section 3.1 that narrative exposure explains idiosyncratic variance in the cross-section, and the evidence in Section 3.3 showing an inverse relationship between narrative exposure and price informativeness.

Finally, a surge in a specific narrative's attention or bias boosts trading volume in stocks exposed to that narrative (Proposition 2-(iii)). This prediction is consistent with the empirical findings on narrative exposures and trading activity in Section 3.4.

In addition, the model predicts that both the proportion of unsophisticated investors and the degree of bias have an adverse impact on price informativeness. We defer the test of this prediction to future research, as it requires a comprehensive analysis of multiple media outlets, an estimation of bias distribution, and an assessment of agent sophistication.

4.2 Model Set-up

Agents and Assets. Time is discrete with a set of infinite periods $\mathcal{T}=\{0,1,2,\dots\}$. There are N risky assets, where each asset $n = 1, \dots, N$ provides per-period dividends

$$D_{n,t} = \bar{D}_n + b_n' f_t + \varepsilon_{n,t}, \tag{7}$$

where $\bar{D}_n$ is a constant term, f_t is a $(K \times 1)$ vector of common factors, b_n is a $(K \times 1)$ vector of factor loadings, and $\varepsilon_{n,t}$ is a residual term with mean zero and variance $\sigma^2_{\varepsilon_n}$. The process $\{f_\tau\}_{\tau\in\mathcal{T}}$ is i.i.d. normal with mean vector $\bar{f}$ and variance matrix Σ_f. Residuals and factors are independent across all leads and lags. Without loss of generality, we assume that risky assets are in zero net supply and set $\bar{f}$ equal to zero.

A continuum of investors trades risky assets each period. To focus on price informativeness, we assume investors are risk-neutral, thereby shutting down any impact of narratives on risk premia. A new cohort of investors is born each period, and investors live for two periods. Investors have zero endowments of risky assets when they enter the economy. In the first period, they trade the N risky securities and a riskless asset with exogenous net return $\bar{r}$. In the second period, investors close all positions, consume, and exit the economy. For ease of exposition, we set $\bar{r}$ and all $\bar{D}_n$'s equal to zero, and we provide the solution to the general case in the Appendix.

We denote with $x_{i,t} = (x_{i,1,t}, \ldots, x_{i,N,t})'$ the $(N \times 1)$ vector of the risky asset holdings of young investor i at time t. Investors incur holding costs equal to $\frac{1}{2} x'_{i,t} C_i x_{i,t}$, where $C_i = \text{diag}(c_{i,1}, \ldots, c_{i,N})$ is a diagonal matrix. Each $c_{i,n}$ captures the investor's preferences and holding costs for each asset in a reduced form.

News and bias. Each period, investors learn about future factor innovations from M news articles published in a media outlet. Each news article focuses on one of L's news topics, or "narratives." We denote the $(L \times 1)$ vector of narratives in period t as z_t. The narratives are related to factor innovations according to the equation:

$$z_t = A f_t + \eta_t, \tag{8}$$

where A is an $(L \times K)$ matrix of constants, f_t is a $K \times 1$ vector of factor innovations, and η_t is an $(L \times 1)$ random vector that is independent of f_t and all other random variables. The process $\{\eta_\tau\}_{\tau \in \mathcal{T}}$ is i.i.d. normal with mean zero and variance matrix Σ_η.

We assume that each news article $m = 1, ..., M$ independently selects one of the L narratives at random according to a probability vector $\theta_t = (\theta_{1,t}, ..., \theta_{L,t})'$, which is independently drawn from the same distribution each period. θ_t determines the probability that each article covers one of the L narratives at time t, so it is the relative narrative attention vector at time t.[14]

[14]Consequently, θ_t is analogous to the narrative attention vector derived from WSJ news articles using the LDA algorithm as described in Section 2.3.

The information content of article m when it selects narrative l at time t is equivalent to the signal

$$s_{m,t} = z_{l,t+1} + \pi_{l,t} + \zeta_{m,t},$$

where $\pi_{l,t}$ is a narrative-specific bias with mean π_l and variance $\pi_l^2\sigma^2$, and the error term $\zeta_{m,t}$ is normally distributed with mean zero and variance M/ω, where ω is a positive constant. The processes $\{\pi_{l,\tau}\}_{\tau\in\mathcal{T}}$ and $\{\zeta_{m,\tau}\}_{\tau\in\mathcal{T}}$ are i.i.d. and independent across narratives and articles.

Thus, the media outlet conveys information to investors that is valuable but biased, and the average values $\pi_1, ..., \pi_L$ capture the persistent components of the media bias.

We consider the limit where $M \uparrow \infty$ and show in Appendix B that the information published by the media outlet is equivalent to the L signals:

$$S_{l,t} = z_{l,t+1} + \pi_{l,t} + \hat{\zeta}_{l,t}; \qquad \text{for } l = 1, ..., L, \tag{9}$$

where $\hat{\zeta}_{l,t} \sim N\left(0, (\omega\theta_{l,t})^{-1}\right)$. Thus, letting $\Theta_t = \text{diag}(\theta_{1,t}\omega, ..., \theta_{L,t}\omega)$, the $(L \times 1)$ vector of signals $S_t = (S_{1,t}, ..., S_{L,t})'$ has precision matrix

$$Var\left(S_t | z_{t+1}, \pi_t\right)^{-1} = \Theta_t, \tag{10}$$

where π_t is the $(L \times 1)$ vector of media biases $\pi_t = (\pi_{1,t}, .., \pi_{L,t})'$. Eq. (10) maps the relative narrative attention θ_t into the precision of investor information. When relative attention to a certain narrative increases, that is, when the corresponding element of θ_t goes up, investors learn more about that narrative from the media outlet.

Investor sophistication. Each investor belongs to one of two classes indexed by R and U. Investors in class R are fully rational and are aware of the media bias in each period, whereas investors in class U are unsophisticated and ignore the media bias. The relative proportion of R and U investors is constant across cohorts. We assume that the structure of the economy is common knowledge and that U investors have dogmatic beliefs.[15] Since all information is public, investor beliefs are the same for all investors in the same class. Thus, we denote the

[15]Therefore, R investors know that U investors have biased beliefs, whereas U investors believe R investors have biased beliefs: $E_i\left(E_j\left(S_{l,t}\right)\right) = 0$ and $E_j\left(E_i\left(S_{l,t}\right)\right) = \pi_l$ for all $i \in R$, $j \in U$, and $l = 1, ..., L$.

expectations of any random variable y as $E_{i,t}(y) = E_{R,t}(y)$ for all $i \in R$ and $E_{i,t}(y) = E_{U,t}(y)$ for all $i \in U$.

Timeline. In each period t, the timeline proceeds as follows. All investors observe realized dividends D_t, factors f_t, and signals S_t. R investors also observe the media bias π_t, while U investors believe S_t is unbiased. Then, young investors submit demand schedules for risky assets based on the information they have observed, and old investors close their positions from the previous period. Market-clearing prices are determined for each risky asset n. Finally, old investors consume and exit the economy.

4.3 Analysis

Prices and returns. It is convenient to express the dividend Eq. (7) in terms of narratives:

$$D_n = \beta_n' z_t + \varphi_{n,t}, \tag{11}$$

where β_n is the $(L \times 1)$ vector of asset-n dividend sensitivities to the L narratives, and $\varphi_{n,t}$ is a residual term that is uncorrelated with f_t and with z_t.[16] Our assumptions regarding investor sophistication imply the following expectations:

$$E_{R,t}\left(D_{n,t+1}\right) = \beta_n' \Phi_t \left(S_t - \pi_t\right); \qquad E_{U,t}\left(D_{n,t+1}\right) = E_{R,t}\left(D_{n,t+1}\right) + \Pi_{n,t}, \tag{12}$$

$$\text{where} \quad \Phi_t = \left(A\Sigma_f A' + \Sigma_\eta\right)\left(A\Sigma_f A' + \Sigma_\eta + \Theta_t^{-1}\right)^{-1} \tag{13}$$

$$\text{and} \quad \Pi_{n,t} = \beta_n' \Phi_t \pi_t. \tag{14}$$

The $(L \times L)$ matrix Φ_t determines the sensitivity of investor narrative expectations to time-t news and depends on the relative attention vector θ_t via the precision matrix Θ_t. The term $\Pi_{n,t}$ is U investor dividend expectation bias for asset n. Intuitively, Eq. (14) shows that $\Pi_{n,t}$ depends on Φ_t times the realized bias π_t, weighted by the asset n's cash flow sensitivities to the L narratives, β_n.

[16]See Eqs. (B3)-(B4) in Appendix B.

Equilibrium. We focus on *linear stationary equilibria* where the price function for each asset $n = 1, \ldots N$ takes the form

$$P_{n,t} = A_n(\theta_t) + B_n(\theta_t)S_t + C_n(\theta_t)\pi_t, \tag{15}$$

where A_n, B_n, and C_n are time-invariant functions.[17] The following proposition solves for equilibrium asset prices and returns. Additionally, it derives the assets' narrative betas that are central to our empirical analysis.

Proposition 1. *(Asset prices and returns) In the unique linear stationary equilibrium:*

(i) *The asset price of security n at time t equals*

$$P_{n,t} = E_{R,t}\left(D_{n,t+1}\right) + \gamma_n \Pi_{n,t}, \tag{16}$$

where $\gamma_n = \frac{\psi_{U,n}}{\psi_{R,n}+\psi_{U,n}}$ and $\psi_{a,n} = \int_a c_{i,n}^{-1} di$, for $a = R, U$.

(ii) *The return $r_{n,t} := P_{n,t} + D_{n,t} - P_{n,t-1}$ equals*

$$r_{n,t} = E_{R,t}\left(D_{n,t+1}\right) + D_{n,t} - E_{R,t-1}\left(D_{n,t}\right) + \gamma_n\left(\Pi_{n,t} - \Pi_{n,t-1}\right). \tag{17}$$

(iii) *Asset-n's beta with respect to narrative l's relative attention, $\beta\left(n,l\right) := \frac{Cov\left(r_{n,t},\theta_{l,t}\right)}{Var\left(\theta_{l,t}\right)}$, equals*

$$\beta\left(n,l\right) = \frac{\gamma_n}{Var\left(\theta_{l,t}\right)} \sum_{j=1}^{L} \pi_j \beta_n' cov\left(\phi_{j,t}, \theta_{l,t}\right), \tag{18}$$

where $\phi_{j,t}$ is the j-th column of the matrix Φ_t.

Proof. See Appendix B. ■

Proposition (1)-(i) shows that an asset price is the rational expectation of the next-period dividend, plus U investors' expectation bias weighted by their trading aggressiveness relative to R investors, γ_n.

[17] In equilibrium, all investors agree on the price function in Eq. (15). Therefore, U investors infer π_t from the price function, attributing it to a bias in R investors' beliefs (see Footnote 15).

Proposition (1)-(ii) decomposes an asset return into a rational and biased-driven part. The rational part includes time-t information about the future dividend and the current dividend's forecast error. The bias-driven part is due to U investors' expectation bias and its evolution over the current and prior periods.

The intuition for Proposition 1-(iii) is as follows. When a narrative receives greater attention, U investors' beliefs load more strongly on that narrative's bias (Eqs. (12)-(14)). U investors have price impact, so the stock return moves in the direction of this narrative's bias, adjusted for cash flow narrative exposures. This mechanism leads to the covariance between narrative attention and stock return in Eq. (18). Notice that without media bias, with investors fully aware of the bias, or if media narrative attention remains constant, all stocks' narrative betas become nil.

Definitions. Next, we introduce two definitions. First, we define price informativeness for asset n as

$$I_n = \frac{Cov\left(D_{n,t+1}, P_{n,t}\right)^2}{Var\left(P_{n,t}\right)}. \tag{19}$$

This definition is standard in market microstructure (e.g., Kacperczyk, Nosal, and Sundaresan, 2022) and is consistent with the approach in Bai, Philippon, and Savov (2016), which forms the basis of our empirical analysis. In our model, Eq. (19) measures the reduction in the posterior dividend uncertainty of an investor who learns from the price using a linear model.[18]

Second, we define the narrative exposure for asset n as in Eq. (3) in Section 2.3:

$$Narrative\ Exposure_n = \frac{\sum_l^L |\beta(n,l)| \times \sigma(\theta_l)}{\sum_l \sigma(\theta_l)}, \tag{21}$$

where each $\beta(n,l)$ is given in Eq. (18).

The next proposition derives the implications of media bias for return volatility, price informativeness, narrative exposure, and trading volume.

[18]Consider the linear model $D_{n,t+1} = a_n + b_n P_{n,t} + e_{n,t+1}$. The variance of $D_{n,t+1}$ conditional on $P_{n,t}$ is the variance of the forecast error $e_{n,t+1}$. Therefore, price information reduces dividend uncertainty by the amount

$$Var\left(D_{n,t+1}\right) - Var\left(e_{n,t+1}\right) = b_n^2 Var\left(P_{n,t}\right) = I_n, \tag{20}$$

where the second equality follows from $b_n = \frac{Cov\left(D_{n,t+1}, P_{n,t}\right)}{Var\left(P_{n,t}\right)}$ and the definition of I_n in Eq. (19). In our empirical analysis, we follow Bai, Philippon, and Savov (2016) and estimate $b_n^2 Var\left(P_{n,t}\right)$ from the cross-section (see Section 3.2).

Proposition 2. *(Narrative exposures, price informativeness, and trading volume)*

(i) *Security n's return variance and price informativeness equal*

$$Var\left(r_{n,t}\right) = SysVar_n + IdVar_n; \qquad I_n = \frac{Var\left[E_{R,t}\left(D_{n,t+1}\right)\right]^2}{Var\left[E_{R,t}\left(D_{n,t+1}\right)\right] + \gamma_n^2 Var(\Pi_{n,t})}, \tag{22}$$

where

$$SysVar_n = \beta_n' \Sigma_f \beta_n; \qquad IdVar_n = \sigma_{\varepsilon_n}^2 + 2\gamma_n^2 Var(\Pi_{n,t}), \tag{23}$$

where $Var\left[E_{R,t}\left(D_{n,t+1}\right)\right]$ is given in Eq. (B15) in Appendix B, and the media-bias-driven component of return volatility $\gamma_n^2 Var(\Pi_{n,t})$ is given by

$$Var(\Pi_{n,t}) = \sum_i^L \sum_j^L \pi_i \pi_j Cov\left(\beta_n' \phi_{i,t}, \beta_n' \phi_{j,t}\right) + \sum_l^L \pi_l^2 \sigma^2 E\left(\beta_n' \phi_{l,t}\right)^2. \tag{24}$$

Therefore, $\gamma_n^2 Var(\Pi_{n,t})$ is positively related to idiosyncratic volatility $IdVar_n$ and negatively related to price informativeness I_n in Eq. (22).

(ii) *The narrative exposure for asset n is related to the media-bias-driven component of return volatility $\gamma_n^2 Var(\Pi_{n,t})$ as follows:*

$$Narrative\ Exposure_n = \sqrt{\gamma_n^2 Var(\Pi_{n,t})} \times \frac{\sum_l^L \left|Corr\left(\Pi_{n,t}, \theta_{l,t}\right)\right|}{\sum_l \sigma(\theta_l)}. \tag{25}$$

(iii) *Trading volume in security n at time t, defined as $TV_{n,t} = \frac{1}{2(\psi_R + \psi_U)} \int_{R \cup U} \mid x_{i,n,t} \mid di$, equals*

$$TV_{n,t} = \gamma(1-\gamma)|\Pi_{n,t} - E(\Pi_{n,t})|. \tag{26}$$

Proof. See Appendix B. ■

Proposition 2-(i) decomposes return volatility into systematic and idiosyncratic parts. The idiosyncratic part depends on idiosyncratic dividend volatility $\sigma_{\varepsilon_n}^2$ and the media-bias-driven component of return volatility $\gamma_n^2 Var(\Pi_{n,t})$.

Proposition 2-(i) also shows that media information has two effects on price informativeness. On the other hand, more precise media information, measured by the variance of the R agents' dividend expectation, brings prices closer to fundamentals, thereby improving price informativeness. On the other hand, media bias decreases price informativeness via the variance of $\gamma_n \Pi_{n,t}$, which measures the price impact of U investors' dividend expectation bias (Eq. 16) and is unrelated to fundamentals.

Proposition 2-(ii) shows how narrative exposure proxies for the media-bias-induced return volatility. This is intuitive because both the $\beta\,(n,l)$'s in Eq. (18) and $\gamma_n^2 Var(\Pi_{n,t})$ in Eq. (24) are driven by media biases weighted by an asset's cash flow narrative exposures times U investors' trading aggressiveness in that asset. Thus, they carry overlapping information. For example, in the case of independent narratives where asset n loads only on narrative l, the proof of the proposition shows that Eq. (25) simplifies to $Narrative\ Exposure_n = \sqrt{\gamma_n^2 Var(\Pi_{n,t})} \times \kappa_l$, where the constant of proportionality κ_l is independent of β_n, γ_n, and π. Therefore, for stocks that load mostly on one narrative, narrative exposures explain most of the cross-sectional variation in the media-bias-induced return volatility and, by extension, in idiosyncratic volatility $IdVar_n$ and price informativeness I_n. For the general case, our empirical analysis in Section 3.1 demonstrates a strong positive cross-sectional relationship between narrative exposures and idiosyncratic variance.

Proposition 2-(iii) shows that a shock to U investors' expectation bias, driven either by an increase in narrative attention or bias, triggers trading among U and R investors due to disagreement about future dividends. The effect is most significant when U and R investors have comparable price impact (i.e., when $\gamma_n = 1/2$).

5 Conclusion

We establish theoretically and empirically that stock return exposure to the shocks in media narrative attention can distort stock prices and decrease their informativeness about future fundamentals. Importantly, we quantify media attention to narratives without measuring their sentiment, that is, in a manner consistent with widely used Natural Language Processing methods, such as Latent Dirichlet Allocation (LDA), for extracting topics from text.

Empirically, the stock prices of firms with high levels of narrative exposure become uninformative about future fundamentals, and price informativeness further deteriorates in periods with high market-wide narrative exposure. While shocks to media narrative attention barely improve the explanatory power of standard factor models, narrative exposure hugely contributes to generating excess volatility in stock returns. Analyzing information channels through which attention to narratives flows to financial markets, we identify narrative exposure as the main characteristic that alone explains 70% to 85% of the cross-sectional variation in idiosyncratic variance and variance due to firm-specific public information. Stocks strongly affected by the narrative attention shocks experience higher average trading volume, indicating that media narrative attention feeds into individual stocks' latent demand.

Using a trading model with time-varying public information production that maps tightly to the LDA methodology employed in our empirical analysis, we demonstrate that in the presence of biased media and investors, media attention to narratives, otherwise uncorrelated with stock returns, affects stock prices. The weight of biased investors in the economy and the level of attention to a particular narrative distorts price informativeness.

Overall, we complement and extend several strands of the literature. Abstracting from predictability and risk premium, we show how media attention to narratives interacts with asset return dynamics, distorting the information content of stock prices with elevated narrative exposure. According to existing studies, media attention to narratives can be useful in predicting returns and defining risk premiums. In contrast, we demonstrate the detrimental media effects on price efficiency—and they are nontrivial, both statistically and economically. Linking to the literature on differences in beliefs, we propose media attention to narratives as a theoretically sound and empirically important source of disagreement in financial markets.

References

Amihud, Y., 2002, "Illiquidity and Stock Returns: Cross-section and Time-series Effects," *Journal of Financial Markets*, 5(1), 31–56.

Bai, J., T. Philippon, and A. Savov, 2016, "Have financial markets become more informative?," *Journal of Financial Economics*, 122(3), 625 – 654.

Baker, S. R., N. Bloom, and S. J. Davis, 2016, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131(4), 1593–1636.

Baloria, V. P., and J. Heese, 2018, "The effects of media slant on firm behavior," *Journal of Financial Economics*, 129(1), 184–202.

Barber, B. M., and T. Odean, 2000, "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *The Journal of Finance*, 55(2), 773–806.

Barberis, N., 2018, "Psychology-based models of asset prices and trading volume," in *Handbook of behavioral economics: applications and foundations 1*. Elsevier, vol. 1, pp. 79–175.

Barberis, N., A. Shleifer, and R. Vishny, 1998, "A Model of Investor Sentiment," *Journal of Financial Economics*, 49(3), 307–343.

Baron, D. P., 2006, "Persistent media bias," *Journal of Public Economics*, 90(1), 1–36.

Blei, D. M., A. Y. Ng, and M. I. Jordan, 2003, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, 3(null), 993–1022.

Bond, P., A. Edmans, and I. Goldstein, 2012, "The real effects of financial markets," *Annu. Rev. Financ. Econ.*, 4(1), 339–360.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer, 2020, "Overreaction in macroeconomic expectations," *American Economic Review*, 110(9), 2748–82.

Brogaard, J., T. H. Nguyen, T. J. Putnins, and E. Wu, 2022, "What Moves Stock Prices? The Roles of News, Noise, and Information," *The Review of Financial Studies*, 35(9), 4341–4386.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu, 2021, "Business News and Business Cycles," Working Paper 29344, National Bureau of Economic Research.

Bybee, L., B. T. Kelly, and Y. Su, 2022, "Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text," working paper 21-09, Johns Hopkins Carey Business School Research Paper.

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.

Cao, J., A. Goyal, S. Ke, and X. Zhan, 2022, "Options trading and stock price informativeness," *Swiss Finance Institute Research Paper*, (19-74).

Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1), 57–82.

Chen, Y., B. Kelly, and W. Wu, 2020, "Sophisticated investors and market efficiency: Evidence from a natural experiment," *Journal of Financial Economics*, 138(2), 316–341.

Cookson, J. A., and M. Niessner, 2020, "Why Don't We Agree? Evidence from a Social Network of Investors," *The Journal of Finance*, 75(1), 173–228.

De Bondt, W. F., and R. Thaler, 1985, "Does the stock market overreact?," *The Journal of Finance*, 40(3), 793–805.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, 1990, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98(4), 703–38.

DellaVigna, S., and E. Kaplan, 2007, "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics*, 122(3), 1187–1234.

Dim, C., K. Koerner, M. Wolski, and S. Zwart, 2022, "Hot off the press: News-implied sovereign default risk," *EIB Working Papers*.

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel, 2020, "Hedging climate change news," *The Review of Financial Studies*, 33(3), 1184–1216.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stock and Bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— , 2015, "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Fama, E. F., and J. D. MacBeth, 1973, "Risk Return and Equilibrium: empirical Tests," *Journal of Financial Political Economy*, 71, 607–636.

Farboodi, M., A. Matray, L. Veldkamp, and V. Venkateswaran, 2021, "Where Has All the Data Gone?," *The Review of Financial Studies*, 35(7), 3101–3138.

Frank, M. Z., and A. Sanati, 2018, "How does the stock market absorb shocks?," *Journal of Financial Economics*, 129(1), 136–153.

Frazzini, A., 2006, "The disposition effect and underreaction to news," *The Journal of Finance*, 61(4), 2017–2046.

Gabaix, X., and R. Koijen, 2021, "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," NBER Working Paper 28967.

Garcia, D., 2013, "Sentiment during recessions," *The Journal of Finance*, 68(3), 1267–1300.

Gentzkow, M., and J. M. Shapiro, 2006, "Media bias and reputation," *Journal of Political Economy*, 114(2), 280–316.

——— , 2010, "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 78(1), 35–71.

Goldman, E., N. Gupta, and R. D. Israelsen, 2021, "Political polarization in financial news," *Available at SSRN 3537841*.

Goldman, E., J. Martel, and J. Schneemeier, 2022, "A theory of financial media," *Journal of Financial Economics*, 145(1), 239–258.

Gurun, U. T. G., and A. W. Butler, 2012, "Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value," *The Journal of Finance*, 67(2), 561–598.

Hanley, K. W., and G. Hoberg, 2019, "Dynamic interpretation of emerging risks in the financial sector," *The Review of Financial Studies*, 32(12), 4543–4603.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh, 2016, "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 119(2), 249–283.

Hillert, A., H. Jacobs, and S. Müller, 2014, "Media makes momentum," *The Review of Financial Studies*, 27(12), 3467–3501.

Kacperczyk, M., S. Sundaresan, and T. Wang, 2020, "Do Foreign Institutional Investors Improve Price Efficiency?," *The Review of Financial Studies*, 34(3), 1317–1367.

Kacperczyk, M. T., J. Nosal, and S. Sundaresan, 2022, "Market power and price informativeness," *Available at SSRN 3137803*.

Koijen, R. S. J., and M. Yogo, 2019, "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127(4), 1475–1515.

Liu, Y., and B. Matthies, 2022, "Long-Run Risk: Is It There?," *The Journal of Finance*.

Manela, A., 2014, "The value of diffusing information," *Journal of Financial Economics*, 111(1), 181–199.

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

Mullainathan, S., and A. Shleifer, 2005, "The market for news," *American Economic Review*, 95(4), 1031–1053.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive-semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Niessner, M., and E. C. So, 2018, "Bad news bearers: The negative tilt of the financial press," *Available at SSRN 3219831*.

Reuter, J., and E. Zitzewitz, 2006, "Do ads influence editors? Advertising and bias in the financial media," *The Quarterly Journal of Economics*, 121(1), 197–227.

Shiller, R. J., 2017, "Narrative Economics," *American Economic Review*, 107(4), 967–1004.

Shiller, R. J., 2020, *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Shleifer, A., and L. H. Summers, 1990, "The noise trader approach to finance," *Journal of Economic Perspectives*, 4(2), 19–33.

Tetlock, P. C., 2007, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance*, 62(3), 1139–1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008, "More than words: Quantifying language to measure firms' fundamentals," *The Journal of Finance*, 63(3), 1437–1467.

Wiesen, T. F., and P. M. Beaumont, 2020, "A Joint Impulse Response Function for Vector Autoregressive Models," *ERN: Other Econometrics: Applied Econometric Modeling in Forecasting (Topic)*.

A Additional Tables

Table A1: Variable Definitions

Variable	Years	Definition
Narrative Exposure		
$\beta^{narr}_{n,t,l}$	1998-2021	Narrative beta for stock n estimated at the end of each year t using daily excess returns, factor realizations and attention shocks to the particular narrative l over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary, CRSP, Own computations.
$Narrative\ Exposure_{n,t}$	1998-2021	Average absolute narrative betas weighted by the standard deviation of attention to each of $l = 1, \ldots, 33$ narratives in a given year: $\frac{\sum_l \lvert\beta^{narr}_{n,t,l}\rvert \times \sigma_t(\theta_l)}{\sum_l \sigma_t(\theta_l)}$. Source: Own computations.
Fundamentals and Stock Characteristics		
$Market\ Cap_{n,t}$	1998-2021	A stock's market capitalization. Source: CRSP.
$Assets_{n,t}$	1998-2021	Total assets (Compustat item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$\ln(Market\ Cap/Assets)_{n,t}$	1998-2021	Log of the ratio of c stock's market capitalization to Total assets. Winsorized annually at 5% and 95%. Source: Compustat NA Annual, CRSP.
$Debt/Assets_{n,t}$	1998-2021	Sum of the book value of long-term debt (Compustat data item DLTT) and the book value of current liabilities (DLC) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Cash/Assets_{n,t}$	1998-2021	Cash and short-term investments (Compustat data item CHE) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$PP\&E/Assets_{n,t}$	1998-2021	Property, plant, and equipment (Compustat data item PPENT) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$EBIT/Assets_{n,t}$	1998-2021	Earnings before interest and taxes (Compustat data item EBIT) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$EBIT_{n,t+h}/Assets_{n,t}$	1998-2021	Earnings before interest and taxes (Compustat data item EBIT) h years from the current year divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Capex/Assets_{n,t}$	1998-2021	Capital expenditures divided by assets. Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$R\&D/Assets_{n,t}$	1998-2021	R&D expenditures (Compustat data item XRD) divided by total assets (Compustat data item AT). Missing values set to zero. Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Turnover_{n,t}$	1998-2021	Turnover relative to shares outstanding. Computed as the daily volume over shares outstanding averaged over all days in a given year. Source: CRSP.
$Illiquidity_{n,t}$	1998-2021	Amihud (2002) Illiquidity measure computed as the daily absolute return over traded volume ratio averaged over all days in a given year (for stocks with at least 63 observations). Winsorized annually at 5% and 95%. Source: CRSP.
$MAX_{n,t}$	1998-2021	The maximum daily return for a stock within each month averaged per year. Source: CRSP.
$Inst.\ Ownership_{n,t}, \%$	1999-2018	Quarterly institutional ownership averaged to the annual level for each year and firm. Source: Thomson Reuters 13F.

Variable	Years	Definition
Factor Betas		
$Market_{n,t}$	1998-2021	Market beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Size\ (SMB)_{n,t}$	1998-2021	Size factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Value\ (HML)_{n,t}$	1998-2021	Value factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Momentum\ (WML)_{n,t}$	1998-2021	Momentum factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Profitability\ (RMW)_{n,t}$	1998-2021	Profitability factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Investment\ (CMA)_{n,t}$	1998-2021	Investment factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
Variance Decomposition Variables		
$IdVar_{n,t}$	1998-2021	Idiosyncratic variance for several factor models (market, three-, four-, and five-factor models) for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 return observations. Computed as the mean-squared error of the fitted model residual. Source: K. French's DataLibrary.
$SysVar_{n,t}$	1998-2021	Systematic variance for several factor models (market, three-, four-, and five-factor models) for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 return observations. Computed as the total variance of daily returns minus the respective idiosyncratic variance. Source: K. French's DataLibrary.
$MktInfo_{n,t}$	1998-2021	Stock variance due to market-wide information. Estimated for each year at the end of December using daily market returns, daily stock signed dollar volume and daily stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.
$PrivateInfo_{n,t}$	1998-2021	Stock variance due firm-specific private information. Estimated for each year at the end of December using daily market returns, daily stock signed dollar volume and daily stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.
$PublicInfo_{n,t}$	1998-2021	Stock variance due to public information. Estimated for each year at the end of December using daily market returns, daily stock signed dollar volume and daily stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.
$Noise_{n,t}$	1998-2021	Stock variance due to noise. Estimated for each year at the end of December using daily market returns, stock signed dollar volume and stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.

B Proofs

Remark In this appendix we solve the model for general D_n's and $\bar{r} > 0$. The expressions in the main text can be recovered by taking the limits as $\bar{r} \to 0$ and $D_n/\bar{r} \to 0$.

Proof of Eq. (9) Consider initially the case where the number of articles M is finite. Each article selects a narrative at random according to the probability vector θ_t. We denote $\mathcal{M}_{l_t}$ the index set of all articles about narrative l at time t and we denote $M_{l,t}$ its cardinality. For each narrative $l = 1, \ldots, L$, the set of signals $\{s_{m,t}\}_{m \in \mathcal{M}_{l,t}}$ is equivalent to the sufficient statistic

$$S_{l,t} = \sum_{m \in \mathcal{M}_{l,t}} \frac{s_{m,t+1}}{M_{l,t}} = z_{l,t} + \pi_{l,t} + \hat{\zeta}_{l,t}; \qquad \text{for } l = 1, ..., L, \tag{B1}$$

where $\hat{\zeta}_{l,t} := \sum_{m \in \mathcal{M}_{l,t}} \frac{\zeta_{m,t}}{M_{l,t}}$. If $M_{l,t} = 0$, then $S_{l,t}$ is pure noise. The precision of $S_{l,t}$ is $Var\left(S_{l,t} | z_{l,t+1}, \pi_{l,t}\right)^{-1} = \frac{M_{l,t}}{M}\omega$. The Law of Large Numbers implies

$$\lim_{M \uparrow \infty} \frac{M_{l,t}}{M} = \theta_{l,t}. \tag{B2}$$

□

Projection of dividends on narratives in Eq. (11) Eqs. (7)-(8) imply the following projection of f_t onto z_t:

$$f_t = \Sigma_f A' \left(A \Sigma_f A' + \Sigma_\eta\right)^{-1} z_t + \nu_t,$$

where ν_t is uncorrelated with z_t. Therefore, the projection of dividends on narratives in Eq. (11) holds for

$$\beta_n' = b_n' \Sigma_f A' \left(A \Sigma_f A' + \Sigma_\eta\right)^{-1} \tag{B3}$$

and

$$\varphi_{n,t} = \epsilon_{n,t} + b_n' \nu_t. \tag{B4}$$

□

Proof of Proposition 1-(i) Our model assumptions imply that young investor i at time t maximizes

$$x_{i,t}' E_{i,t} \left(P_{t+1} + D_{t+1} - (1 + \bar{r}) P_t\right) - \frac{1}{2} x_{i,t}' C_i x_{i,t} \tag{B5}$$

where D_t and P_t denote, respectively, the $(N \times 1)$ vectors of asset dividends and prices at time t, and $E_{i,t}$ denotes the time-t conditional expectation of investor i. The solution to (B5) gives, asset by asset,

$$x_{i,n,t} = c_{i,n}^{-1} E_{i,t} \left(P_{n,t+1} + D_{n,t+1} - (1+\bar{r}) P_{n,t} \right). \tag{B6}$$

Market clearing for asset n requires

$$\int\limits_R x_{i,n,t} di + \int\limits_U x_{i,n,t} di = 0.$$

Substituting Eq. (B6) in the market clearing condition and using the definitions of $\psi_{R,n}, \psi_{U,n}$ and γ_n in Proposition 1-(i), we can solve for the market clearing price $P_{n,t}$ as

$$P_{n,t} = \frac{(1-\gamma_n) E_{R,t} (P_{n,t+1}) + \gamma_n E_{U,t} (P_{n,t+1}) + (1-\gamma_n) E_{R,t} (D_{n,t+1}) + \gamma_n E_{U,t} (D_{n,t+1})}{(1+\bar{r})}. \tag{B7}$$

Using the conjectured price function in Eq. (15) and the fact that $E_{R,t}(S_{t+1}) = \pi$ and $E_{R,t}(S_{t+1}) = 0$, we obtain

$$(1-\gamma_n) E_{R,t} (P_{n,t+1}) + \gamma_n E_{U,t} (P_{n,t+1}) = \bar{A}_n + \left[(1-\gamma_n) \bar{B}_n + \bar{C}_n \right] \pi, \tag{B8}$$

where $\bar{A}_n = E_i [A_n (\theta_{t+1})]$ and $\bar{B}_n = E_i [B_n (\theta_{t+1})]$ and $\bar{C}_n = E_i [C_n (\theta_{t+1})]$. Using Eq. (12) we obtain

$$(1-\gamma^n) E_{R,t} (D_{n,t+1}) + \gamma^n E_{U,t} (D_{n,t+1}) = \bar{D}_n + \beta_n' \Phi_t S_t - (1-\gamma^n) \beta_n' \Phi_t \pi_t. \tag{B9}$$

Substituting Eqs. (B8)-(B9) into Eq. (B7) and matching coefficients with Eq. (15) we find

$$A_n(\theta_t) = \frac{D_n}{\bar{r}}; \quad B_n(\theta_t) = \frac{\beta_n' \Phi_t}{1+\bar{r}}; \quad C_n(\theta_t) = -(1-\gamma_n) B_n (\theta_t).$$

Therefore, we obtain

$$P_{n,t} = \frac{D_n}{\bar{r}} + \frac{\beta_n' \Phi_t S_t - (1-\gamma_n) \beta_n' \Phi_t \pi_t}{1+\bar{r}}.$$

Rearranging terms and using Eqs. (12)-(14) yields

$$P_{n,t} = \frac{\bar{D}_n / \bar{r} + E_{R,t} (D_{n,t+1}) + \gamma_n \Pi_{n,t}}{1+\bar{r}} \tag{B10}$$

$\square$

Proof of Proposition 1-(ii) Eq. (B10) and the return definition in the text immediately imply

$$r_{n,t} = \frac{E_{R,t}\left(D_{n,t+1}\right) - \bar{D}_n}{1+\bar{r}} + D_{n,t} - E_{R,t-1}\left(D_{n,t}\right) + \gamma_n \frac{\Pi_{n,t} - (1+\bar{r})\Pi_{n,t-1}}{1+\bar{r}}. \tag{B11}$$

$\square$

Proof of Proposition 1-(iii) Using Eq. (B11) we compute:

$$Cov\left(r_{n,t}, \theta_{l,t}\right) = \frac{Cov\left(E_{R,t}\left(D_{n,t+1}\right), \theta_{l,t}\right) + \gamma_n Cov\left(\Pi_{n,t}, \theta_{l,t}\right)}{1+\bar{r}}.$$

We can write

$$Cov\left(E_{R,t}\left(D_{n,t+1}\right), \theta_{l,t}\right) = Cov\left(D_{n,t+1}, \theta_{l,t}\right) + Cov\left(E_{R,t}\left(D_{n,t+1}\right) - D_{n,t+1}, \theta_{l,t}\right).$$

Since $D_{n,t+1}$ and $\theta_{l,t}$ are independent and the expectation error $E_{R,t}\left(D_{n,t+1}\right) - D_{n,t+1}$ is orthogonal to time-t information, we conclude that $Cov\left(E_{R,t}\left(D_{n,t+1}\right), \theta_{l,t}\right) = 0$, and, therefore,

$$\beta(n,l) = \frac{\gamma_n}{(1+\bar{r})Var(\theta_{l,t})} Cov\left(\Pi_{n,t}, \theta_{l,t}\right). \tag{B12}$$

Next, we observe that

$$\begin{aligned}
\Pi_{n,t} &= \beta_n' \Phi_t \pi_t \\
&= vec\left(\beta_n' \Phi_t \pi_t\right) \\
&= \left(\pi_t' \otimes \beta_n'\right) vec(\Phi_t) \\
&= \left(\pi_{1,t}\beta_{n,t}', ..., \pi_{L,t}\beta_n'\right)\left(\phi_{1,t}', ..., \phi_{L,t}'\right)' \\
&= \sum_{l=1}^{L} \pi_{l,t}\beta_n' \phi_{l,t}, \qquad \text{(B13)}
\end{aligned}$$

where $\phi_{l,t}$ denotes the l-th column of Φ_t. Finally, we compute

$$\begin{aligned} Cov\left(\Pi_{n,t},\theta_{l,t}\right) &= \sum_{j=1}^{L} Cov\left(\pi_{j,t}\beta_n'\phi_{j,t},\theta_{l,t}\right) \\ &= \sum_{j=1}^{L} E\left[\pi_{j,t}Cov\left(\beta_n'\phi_{j,t},\theta_{l,t}\right)\right] + \sum_{j=1}^{L} Cov\left[\pi_{j,t}E\left(\beta_n'\phi_{j,t}\right),E(\theta_{l,t})\right] \\ &= \sum_{j=1}^{L} \pi_j Cov\left(\beta_n'\phi_{j,t},\theta_{l,t}\right), \end{aligned}$$

where the first equality follows from Eq. (B13) in the previous derivation and the second one from the law of total covariance. Plugging the last expression into Eq. (B12) yields Eq. (18).

□

Proof of Proposition 2-(i) First, we prove that $Var\left(r_{n,t}\right) = SysVar_n + IdVar_n$ in Eq. (22). Using Eq. (17) we compute:

$$\begin{aligned} Var\left(r_{n,t}\right) \quad = \quad & \frac{Var\left(E_{R,t}\left(D_{n,t+1}\right)\right)}{\left(1+\bar{r}\right)^2} + Var\left(D_{n,t} - E_{R,t-1}\left(D_{n,t}\right)\right) + \gamma_n^2 \frac{Var\left(\Pi_{n,t} - \left(1+\bar{r}\right)\Pi_{n,t-1}\right)}{\left(1+\bar{r}\right)^2} \\ & +2\frac{Cov\left(E_{R,t}\left(D_{n,t+1}\right), D_{n,t} - E_{R,t-1}\left(D_{n,t}\right)\right)}{1+\bar{r}} \\ & +2\gamma_n \frac{Cov\left(D_{n,t} - E_{R,t-1}\left(D_{n,t}\right), \Pi_{n,t} - \left(1+\bar{r}\right)\Pi_{n,t-1}\right)}{1+\bar{r}} \\ & +2\gamma_n \frac{Cov\left(E_{R,t}\left(D_{n,t+1}\right), \Pi_{n,t} - \left(1+\bar{r}\right)\Pi_{n,t-1}\right)}{\left(1+\bar{r}\right)^2} \end{aligned}$$

Our assumptions imply that the conditional expectation $E_{R,t}\left(D_{n,t+1}\right)$ is independent over time and independent of $D_{n,t}$. This implies that the first covariance term in the above expression equals zero. Next, we show that $\Pi_{n,t}$ is uncorrelated with $E_{R,t}\left(D_{n,t+1}\right)$. We have:

$$Cov\left(E_{R,t}\left(D_{n,t+1}\right),\Pi_{n,t}\right) = Cov\left(D_{n,t+1},\Pi_{n,t}\right) + Cov\left(E_{R,t}\left(D_{n,t+1}\right) - D_{n,t+1},\Pi_{n,t}\right). \qquad \text{(B14)}$$

$\Pi_{n,t}$ is a function of θ_t and π_t, which are independent of $D_{n,t+1}$. Thus, the first covariance term in the r.h.s. of Eq. (B14) is zero. Since the expectation error $E_{R,t}\left(D_{n,t+1}\right) - D_{n,t+1}$ is orthogonal to time-t information, also the second term in Eq. (B14) is zero. Furthermore, $\Pi_{n,t}$ is independent over time and independent of $D_{n,t}$ across all leads and lags. This, together with Eq. (B14), implies that also the second and third covariance terms in the expression for $Var(r_{n,t})$ are equal to zero.

Next, the law of total variance implies

$$Var\left(E_{R,t}\left(D_{n,t+1}\right)\right) = Var\left(D_{n,t+1}\right) - E\left[Var_{R,t}\left(D_{n,t+1}\right)\right],$$

and the standard conditional variance formula for normally distributed random variables gives

$$\begin{aligned} Var_{R,t}\left(D_{n,t+1}\right) &= Var_R\left(D_{n,t+1}\right) - Cov_R\left(D_{n,t+1}, S_t\right) Var_R\left(S_t\right)^{-1} Cov_R\left(D_{n,t+1}, S_t\right)' \\ &= \beta_n' \Sigma_f \beta_n + \sigma_{\varepsilon,n}^2 - \beta_n' \left(A\Sigma_f A' + \Sigma_\eta\right)\left(A\Sigma_f A' + \Sigma_\eta + \Theta_t^{-1}\right)^{-1}\left(A\Sigma_f A' + \Sigma_\eta\right)\beta_n. \end{aligned}$$

Therefore,

$$Var_{R,t}\left(E_{R,t}\left(D_{n,t+1}\right)\right) = \beta_n' \left(A\Sigma_f A' + \Sigma_\eta\right) E\left[\left(A\Sigma_f A' + \Sigma_\eta + \Theta_t^{-1}\right)^{-1}\right]\left(A\Sigma_f A' + \Sigma_\eta\right)\beta_n. \tag{B15}$$

Next, we note that the variance of the forecast error

$$Var\left(D_{n,t} - E_{R,t-1}\left(D_{n,t}\right)\right) = Var\left(D_{n,t}\right) - Var\left(E_{R,t-1}\left(D_{n,t}\right)\right).$$

Summing up terms and using stationarity, we obtain

$$Var\left(r_{n,t}\right) = \beta_n' \Sigma_f \beta_n + \frac{1-(1+\bar{r})^2}{(1+\bar{r})^2} Var\left(E_{R,t}\left(D_{n,t+1}\right)\right) + \sigma_{\varepsilon,n}^2 + \gamma_n^2 \frac{1+(1+\bar{r})^2}{(1+\bar{r})^2} Var(\Pi_{n,t})$$

Finally, we compute

$$\begin{aligned} Var\left(\Pi_{n,t}\right) &= Var\left(\sum_{j=1}^{L} \pi_{j,t}\beta_n' \phi_{j,t}\right) \\ &= \sum_{j=1}^{L}\sum_{i=j}^{L} E\left[\pi_{i,t}\pi_{j,t} Cov\left(\beta_n'\phi_{i,t}, \beta_n'\phi_{j,t}\right)\right] + Var\left[\sum_{j=1}^{L}\pi_{j,t}E(\beta_n'\phi_{j,t})\right] \\ &= \sum_{j=1}^{L}\sum_{i=j}^{L} \pi_i\pi_i Cov\left(\beta_n'\phi_{i,t}, \beta_n'\phi_{j,t}\right) + \sum_{j=1}^{L}\pi_j^2\sigma^2 E(\beta_n'\phi_{j,t})^2, \end{aligned}$$

where the second equality follows from the law of total variance and the third from the independence of biases across narratives.

Next, we prove the formula for I_n in Eq. (22). Using the formula for $P_{n,t}$ in Eq. (B10) and the fact that $E_{R,t}\left(D_{n,t+1}\right)$ and $\Pi_{n,t}$ are uncorrelated, we compute

$$Var\left(P_{n,t}\right) = \frac{Var\left(E_{R,t}\left(D_{n,t+1}\right)\right) + \gamma_n^2 Var\left(\Pi_{n,t}\right)}{\left(1+\bar{r}\right)^2}.$$

Next, using the formula for $P_{n,t}$ in Eq. (B10) and the fact that $D_{n,t+1}$ and $\Pi_{n,t}$ are uncorrelated, we compute

$$Cov\left(D_{n,t+1}, P_t\right) = \frac{Cov\left(D_{n,t+1}, E_{R,t}\left(D_{n,t+1}\right)\right)}{1+\bar{r}} = \frac{Var\left(E_{R,t-1}\left(D_{n,t}\right)\right)}{1+\bar{r}}.$$

Therefore,

$$I_n = \frac{\left[Var\left(E_{R,t-1}\left(D_{n,t}\right)\right)\right]^2}{Var\left(E_{R,t}\left(D_{n,t+1}\right)\right) + \gamma_n^2 Var\left(\Pi_{n,t}\right)}.$$

□

Proof of Proposition 2-(ii) Using Eq. (B11) into Eq. (B6) we have, for $i \in R$,

$$\begin{aligned} x_{i,n,t} &= c_{i,n}^{-1} E_{R,t}\left[\frac{E_{R,t+1}\left(D_{n,t+2}\right) - \bar{D}_n}{1+\bar{r}} + D_{n,t+1} - E_{R,t}\left(D_{n,t+1}\right) + \gamma_n \frac{\Pi_{n,t+1} - (1+\bar{r})\Pi_{n,t}}{1+\bar{r}}\right] \\ &= c_{i,n}^{-1}\gamma_n \frac{E_{R,t}\left(\Pi_{n,t+1}\right) - (1+\bar{r})\Pi_{n,t}}{1+\bar{r}}. \end{aligned}$$

Aggregating across agents and using the market clearing condition gives

$$TV_{n,t} = \gamma(1-\gamma)|\Pi_{n,t} - \frac{E(\Pi_{n,t})}{1+\bar{r}}|. \tag{B16}$$

□

Proof of Proposition 2-(iii) We have:

$$Corr\left(\Pi_{n,t}, \theta_{l,t}\right) = \frac{Cov\left(\gamma_n \Pi_{n,t}, \theta_{l,t}\right)}{\sqrt{Var\left(\gamma_n \Pi_{n,t}\right) Var\left(\theta_{l,t}\right)}} = \beta\left(n,l\right) \frac{\sqrt{Var\left(\theta_{l,t}\right)}}{\sqrt{Var\left(\gamma_n \Pi_{n,t}\right)}},$$

where the first equality follows from the definition of correlation and the second from Eq. (18). Rearranging terms and taking the absolute value gives

$$|\beta\left(n,l\right)|\sqrt{Var\left(\theta_{l,t}\right)} = \sqrt{Var\left(\gamma_n \Pi_{n,t}\right)} \times |Corr\left(\Pi_{n,t},\theta_{l,t}\right)|,$$

Summing over l, dividing by $\sum_l \sigma(\theta_l)$, and using the definition of narrative exposure in Eq. (21) by gives Eq. (25).

When asset n loads only on narrative l and narratives are independent, we have

$$\begin{aligned}
\frac{\sum_l^L |Corr\left(\Pi_{n,t},\theta_{l,t}\right)|}{\sum_l \sigma(\theta_l)} &= \frac{\sum_l^L |Corr\left(\beta_{n,l}\pi_l \frac{\omega\theta_{l,t}}{\tau_{z_l}+\omega\theta_{l,t}},\theta_{j,t}\right)|}{\sum_l \sigma(\theta_l)} \\
&= \frac{\sum_l^L |Corr\left(\frac{\omega\theta_{l,t}}{\tau_{z_l}+\omega\theta_{l,t}},\theta_{j,t}\right)|}{\sum_l \sigma(\theta_l)} \\
&= \kappa_l,
\end{aligned}$$

which is independent of β_n, γ_n, and π. □

Online Appendix

to

"Media Narratives and Price Informativeness"

Version: July 27, 2023

OA Data Processing and Construction of Variables

OA.1 News Text Processing

We provide a brief summary of the LDA algorithm and refer interested readers to the original paper (Blei, Ng, and Jordan, 2003) for a detailed description. LDA gives text a hierarchical structure, where documents (news articles) are composed of topical narratives containing words. Precisely, each document has a probability distribution over latent narratives, with parameter $\alpha > 0$, and each narrative is defined by a probability distribution over words with parameter $\beta > 0$. α controls the sparsity of narratives in a document, while β controls the sparsity of words in a narrative. LDA treats a document as a mixture of narratives and a narrative as a mixture of words, such that documents overlap each other rather than being separated into discrete groups.

Training the LDA algorithm boils down to finding the optimal number of latent narratives L that best fit the data. Fitting the LDA algorithm on a corpus of documents with a chosen L yields two outputs: the distribution of word frequencies for each narrative, and the distribution of narratives across documents. For each document, the narrative distribution is a vector of loadings that reflect how much attention is devoted to each narrative in the document, such that higher loading for a particular narrative indicates that the document is more likely associated with that narrative.

We train the LDA algorithm using standard cross-validation and grid search procedures. We first convert the processed text corpus into a document term matrix whose rows are the news articles and columns the unique single words (unigrams) and two-word combinations (bigrams) in the text corpus, excluding terms that occur in less than 0.5% of the text corpus to reduce noise. These unigrams and bigrams constitute the feature space for grouping articles into topical narratives. Next, We use each article's WSJ section name and year of appearance in the WSJ archive as a group variable to split the text corpus into five equal train-test folds for cross-validation. This allows us maintain similar proportion of articles in each section each year throughout training and validation samples. Finally, we search for the number of narratives, L, that minimizes (maximizes) the average test set perplexity (log-likelihood) score.

Figure OA1 summarizes the WSJ news text corpus and our machine learning model training. Panel A shows, on the left axis, the monthly number of news articles in our WSJ historical web archive, and, on the right axis, the number of words in these articles. We observe substantial variations in both the volume of publications and the length of publications over time. Panel B depicts the convergence of the average test set log-likelihood (in millions) to its maximum

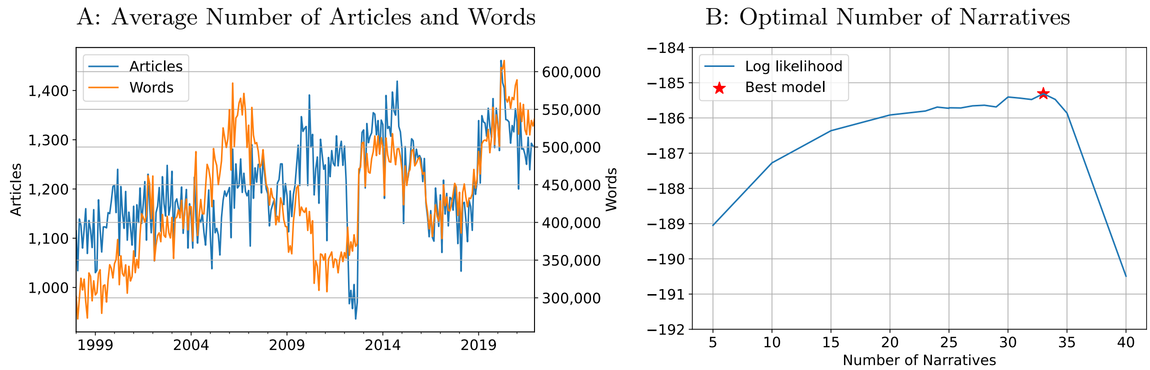

Figure OA1: Article Counts and Model Training. In Panel A, the figures show the total number of articles in our WSJ news corpus per month (left y-axis) and the total number of words in those articles per month (right y-axis) after our preprocessing procedure. Panel B depicts the number of topical narratives in the LDA model that best characterize our news corpus.

across the number of narratives, L, during the LDA model training. The figure indicates that 33 topical narratives optimally characterize our WSJ text corpus.

OA.2 Information Channels via Variance Decomposition

We obtain the information channels affecting stock returns by two different methods of variance decomposition. We perform estimation separately for each firm and each year using daily returns and factor realizations within the year. First, we estimate several linear factor models of the form

$$r_{n,\tau} = \alpha_{n,t} + \beta_{n,t}^{\top} F_{\tau} + \varepsilon_{n,\tau}, \tag{OA17}$$

where $r_{n,\tau}$ is stock n's excess return on day τ in year t, F is the vector of factor realizations on day τ. We use the market model, the three-factor Fama and French (1993) model, the four-factor Carhart (1997) model, and the five-factor Fama and French (2015) model. After estimating each model for firm n in year t we compute the idiosyncratic variance $IdVar_{n,t}$ as the mean-squared error of the residuals, and the systematic variance $SysVar_{n,t}$ as the total variance minus idiosyncratic variance.

Second, we decompose the total stock return variance following the procedure outlined in "Appendix A: Estimation of the structural VAR" in Brogaard, Nguyen, Putnins, and Wu (2022). For the full procedure, we refer our readers to the original paper. Below we outline the major steps of the procedure (freely copying some parts of the original paper) and specific decisions we made in our analysis. The stock return is decomposed into the following parts:

$$r_\tau = \underbrace{\mu}_{\text{discount rate}} + \underbrace{\theta_{r_m}\varepsilon_{r_m,\tau}}_{\text{market-wide info}} + \underbrace{\theta_x\varepsilon_{x,\tau}}_{\text{private info}} + \underbrace{\theta_r\varepsilon_{r,\tau}}_{\text{public info}} + \underbrace{\Delta s_\tau}_{\text{noise}}, \tag{OA18}$$

where $\varepsilon_{r_m,\tau}$ is the unexpected innovation in the market return and $\theta_{r_m}\varepsilon_{r_m,\tau}$ is the market-wide information incorporated into stock prices, $\varepsilon_{x,\tau}$ is an unexpected innovation in signed dollar volume and $\theta_x\varepsilon_{x,\tau}$ is the firm-specific information revealed through trading on private information, and $\varepsilon_{r,\tau}$ is the innovation in the stock price producing the $\theta_r\varepsilon_{r,\tau}$ that is the remaining part of firm-specific information not captured by trading on private information. The components above are obtained from a structural vector autoregression (VAR) model with five lags estimated for market returns $r_{m,\tau}$, signed dollar volume of trading in the given stock x_τ, and stock returns r_τ:

$$\begin{aligned} r_{m,\tau} &= \sum_{l=1}^{5} a_{1,l} r_{m,\tau-l} + \sum_{l=1}^{5} a_{2,l} x_{\tau-l} + \sum_{l=1}^{5} a_{3,l} r_{\tau-l} + \varepsilon_{r_m,\tau} \\ x_\tau &= \sum_{l=0}^{5} b_{1,l} r_{m,\tau-l} + \sum_{l=1}^{5} b_{2,l} x_{\tau-l} + \sum_{l=1}^{5} b_{3,l} r_{\tau-l} + \varepsilon_{x,\tau} \\ r_\tau &= \sum_{l=0}^{5} c_{1,l} r_{m,\tau-l} + \sum_{l=1}^{5} c_{2,l} x_{\tau-l} + \sum_{l=1}^{5} c_{3,l} r_{\tau-l} + \varepsilon_{r,\tau} \end{aligned} \tag{OA19}$$

The required parameters are obtained by first estimating a reduced-form VAR

$$\begin{aligned} r_{m,\tau} &= a_0^* + \sum_{l=1}^{5} a_{1,l}^* r_{m,\tau-l} + \sum_{l=1}^{5} a_{2,l}^* x_{\tau-l} + \sum_{l=1}^{5} a_{3,l}^* r_{\tau-l} + e_{r_m,\tau} \\ x_\tau &= b_0^* + \sum_{l=1}^{5} b_{1,l}^* r_{m,\tau-l} + \sum_{l=1}^{5} b_{2,l}^* x_{\tau-l} + \sum_{l=1}^{5} b_{3,l}^* r_{\tau-l} + e_{x,\tau} \\ r_\tau &= c_0^* + \sum_{l=1}^{5} c_{1,l}^* r_{m,\tau-l} + \sum_{l=1}^{5} c_{2,l}^* x_{\tau-l} + \sum_{l=1}^{5} c_{3,l}^* r_{\tau-l} + e_{r,\tau} \end{aligned} \tag{OA20}$$

and then using the reduced form error covariances to recover the structural VAR parameters, including variances of the residuals $\sigma_{r_m}^2$, σ_x^2, and σ_r^2.

Parameters θ_{r_m}, θ_x, θ_r are defined as the long-run cumulative return response functions in the structural model and are computed by feeding through the reduced model the equivalent reduced form shocks. We use for this purpose the joint impulse response function derived in Wiesen and Beaumont (2020).

The variance components are then computed as follows:

$$MktInfo = \theta_{r_m}\sigma_{r_m}^2,\ PrivateInfo = \theta_x\sigma_x^2,\ PublicInfo = \theta_r\sigma_r^2, \tag{OA21}$$

$$Noise = Total\ Variance - MktInfo - PrivateInfo - PublicInfo.$$

OB Additional Tables

	$Narrative\ Exposure_{n,t}$	$IdVar_{n,t}$	$SysVar_{n,t}$	$MktInfo_{n,t}$	$PrivateInfo_{n,t}$	$PublicInfo_{n,t}$	$Noise_{n,t}$
Panel A: Narrative exposure.							
$Narrative\ Exposure_{n,t}$	1.000	0.885	0.214	0.432	0.753	0.823	0.748
Panel B: Variance decomposition.							
$IdVar_{n,t}$	0.885	1.000	0.136	0.421	0.768	0.904	0.882
$SysVar_{n,t}$	0.214	0.136	1.000	0.642	0.168	0.182	0.062
$MktInfo_{n,t}$	0.432	0.421	0.642	1.000	0.409	0.499	0.254
$PrivateInfo_{n,t}$	0.753	0.768	0.168	0.409	1.000	0.726	0.523
$PublicInfo_{n,t}$	0.823	0.904	0.182	0.499	0.726	1.000	0.687
$Noise_{n,t}$	0.748	0.882	0.062	0.254	0.523	0.687	1.000
Panel D: Factor model betas.							
$Market\ Beta_{n,t}$	-0.082	-0.081	0.442	0.276	-0.002	-0.047	-0.119
$Size\ (SMB)\ Beta_{n,t}$	0.195	0.154	0.351	0.269	0.196	0.176	0.070
$Value\ (HML)\ Beta_{n,t}$	0.013	0.055	-0.027	0.027	0.019	0.070	0.076
$Mom.\ (WML)\ Beta_{n,t}$	-0.167	-0.180	-0.137	-0.154	-0.152	-0.190	-0.153
Panel D: Fundamentals and market characteristics.							
$\ln(Assets)_{n,t}$	-0.538	-0.499	0.124	-0.116	-0.399	-0.459	-0.443
$\ln(Market\ Cap/Assets)_{n,t}$	-0.073	-0.086	0.092	0.058	-0.022	-0.085	-0.145
$EBIT_{n,t}/Assets_{n,t}$	-0.542	-0.484	-0.059	-0.230	-0.456	-0.485	-0.373
$Debt_{n,t}/Assets_{n,t}$	-0.012	-0.000	0.015	-0.021	-0.012	0.006	0.026
$Cash_{n,t}/Assets_{n,t}$	0.206	0.129	0.099	0.131	0.172	0.128	0.039
$PP\&E_{n,t}/Assets_{n,t}$	-0.111	-0.075	-0.012	-0.069	-0.088	-0.079	-0.037
$Sales_{n,t}/Assets_{n,t}$	-0.060	-0.025	-0.106	-0.080	-0.060	-0.037	0.028
$Capex_{n,t}/Assets_{n,t}$	-0.008	0.031	0.039	0.031	0.027	0.039	0.026
$R\&D_{n,t}/Assets_{n,t}$	0.302	0.232	0.071	0.150	0.267	0.232	0.133
$Turnover_{n,t}$	0.226	0.116	0.304	0.236	0.251	0.111	0.004
$Illiquidity_{n,t}$	0.281	0.381	-0.185	-0.046	0.153	0.295	0.487
$MAX_{n,t}$	0.737	0.635	0.169	0.307	0.595	0.602	0.519

Table OB1: Correlations for Selected Variables.
The table shows the unconditional correlations among selected variables computed from the firm-year panel data. Each year, all continuous variables are winsorized at 5% and 95% levels.

Table OB2: Most Relevant Terms for the Identified Narratives.
This table shows the top 50 terms with the largest rescaled weights for each narrative and groups (i.e., broader themes) manually assigned to narratives based on the similarity of the top terms.

Meta Theme	Regulation		Macroeconomy				Equity Markets				
Narrative	1	2	3	4	5	6	7	8	9	10	11
1	request comment	court	unemployment	state own	central bank	fed	street journal	join conversation	common stock	franc	third quarter
2	plan would	justice	gdp	renewable	tariff	monetary policy	dow	nasdaq composite	value million	samsung	cent share
3	trade deal	attorney	labor market	follower	ecb	euro zone	jones	composite index	com complete	thomson	net income
4	commission say	white house	unemployment rate	world bank	imf	fannie	dow jones	index rise	million cash	reuters	million cent
5	state legislature	prosecutor	labor department	life insurance	finance minister	freddie	newswires	full list	regulatory filing	pence	quarter end
6	deny wrongdoing	criminal	consumer price	corporate governance	international monetary	raise rate	jones newswires	blue chip	amortization	ftse	quarter net
7	federal state	supreme	jobless	bnp paribas	monetary fund	economist say	stoxx	index fall	capital partners	nobel	revenue rise
8	rule would	supreme court	consumer spending	privatization	ministry say	treasury note	stoxx index	technology stock	deal value	samsung electronics	quarter profit
9	rule require	justice department	home sale	public private	large economy	investment officer	dow industrial	index gain	thomson financial	financial times	quarter result
10	clause	allege	say economist	economic development	trading partner	rate rise	online journal	markets data	buy stake	cac	share compare
11	would require	allegation	statistics	subsidy	central banker	mae	dutch	go markets	earning interest	franc franc	quarter revenue
12	regulate	attorney general	global growth	taxpayer	policy meeting	sarah	jones stoxx	center wsjmarkets	early company	write sarah	income million
13	school district	law enforcement	annual rate	city state	finance ministry	home price	royal dutch	wsjmarkets com	depreciation amortization	swiss	early period
14	proposal would	guilty	trade deficit	welfare	late dollar	term interest	dollar term	wsjmarkets	gs	abn	earning cent
15	appeals	cuomo	consumer confidence	state government	dollar yen	fannie mae	world stock	index close	interest taxis	market committee	analyst expectation
16	law require	jury	seasonally adjust	poverty	late euro	freddie mac	jones global	track stock	million stock	scotland	operating profit
17	deadline	allegedly	headwind	budget	yuan	fed chairman	jones world	decliners	share outstanding	cac index	report net
18	court ruling	alleged	department report	per year	trade organization	bear market	indexes	price large	expect close	federal open	thomson reuters
19	bargaining	impeachment	job growth	sand	foreign exchange	bull market	newswires editor	gainers	stock new	amro	quarter loss
20	trade agreement	testify	previous month	private sector	monetary	fed official	sector dow	close high	cash stock	abn amro	revenue fall
21	union say	plead	industrial production	taxis	foreign currency	million euro	global indexes	point nasdaq	value company	swiss franc	quarter company
22	say federal	administration official	jobless claim	charter	communist party	cut interest	wall street	major index	ipo	ftse index	first call
23	proposal	convict	straight month	insurance company	ese	say economy	est	stock average	stake company	amsterdam	say net
24	federal agency	defendant	manufacturing sector	entitlement	economic crisis	rate increase	wall	index lose	company board	penny	report earning
25	new law	plead guilty	survey thomson	tax	foreign investment	low interest	journal	index add	public offering	ing	current quarter
26	regulation	qaeda	seasonally	exceptional	exchange rate	interest rate	laggard	tech stock	deal expect	france	analyst poll
27	authorization	plaintiff	month ahead	bnp	currency	raise interest	street	exchange close	initial public	deutsche bank	analyst survey
28	would force	indictment	create job	debt crisis	competitiveness	rate fall	exxon mobil	index end	stock option	korea	quarter compare
29	times report	federal court	datum release	federal government	global financial	rate would	asian	close see	sell stake	deutsche	share earning
30	comply	subpoena	commerce department	offshore	global economy	cut rate	edt	stock end	data center	sterling	quarter say
31	compliance	clinton	wage	rebuild	rest world	economist expect	poor stock	stock gain	go public	british	quarter sale
32	chamber commerce	prosecution	unemployed	infrastructure	economic policy	slow pace	three quarter	index climb	company billion	paris	fiscal first
33	would also	wrongdoing	durable	waste	geopolitical	chief investment	tim	benchmark index	regulatory approval	insurer	revenue increase
34	would allow	judicial	employment	percent	foreign investor	market go	rocky	point close	blackstone	seoul	fiscal fourth
35	eu	federal judge	report show	project	strong dollar	rate cut	index track	chip stock	public company	london	early revenue
36	federal regulator	lawyer say	slow growth	program say	global economic	low rate	flock	share change	agree sell	open market	quarter up
37	unanimously	federal prosecutor	fast pace	casino	imbalance	rate low	jones industrial	index slip	billion cash	nv	fiscal third
38	local government	misconduct	median	large part	economic recovery	bull	quarterly earning	technology share	trade company	wind down	quarter million
39	free trade	overturn	payroll	corrupt	easing	job market	times	profit taking	ipos	pretax profit	fiscal second
40	state federal	verdict	growth slow	scarce	central	rate say	milestone	bovespa	closing price	pound	expect report
41	would receive	judge say	contraction	corruption	election campaign	many investor	industrial average	stock advance	new company	consortium	say earning
42	uniform	uphold	growth rate	government plan	bilateral	quantitative	europe	nasdaq index	company raise	plc	earning share
43	audit	dissent	purchasing	sustainable	dollar trade	year date	also provide	issue lead	agree buy	suisse	previous quarter
44	eligible	inspector general	labor	funding	government spending	risky	rattle	volume billion	shareholder	credit suisse	post net
45	would seek	federal bureau	hourly	wealthy	treasury secretary	curve	reassure	general index	filing	ab	poll thomson
46	levy	bureau investigation	minimum wage	incentive	policy maker	investment strategist	news	ge	spinoff	electronics	quarter last
47	exemption	appeal court	many economist	write joseph	say speech	selloff	overweight	small cap	combined	foreign company	earning estimate
48	adhere	statute	labor cost	public policy	grim	market could	think people	index drop	financial statement	suisse first	earning forecast
49	would need	prosecute	chief economist	move ahead	intervention	short term	euro	close low	company own	press report	report third
50	submit	district attorney	economist	initiative	economy say	relatively low	consumer electronic	gain point	holdings inc	hsbc	report second

Continues

Meta Theme	Politics	Fixed Income		Energy Markets	Consumer Staples	Healthcare	Automotive	Telecoms & Social Media			Entertainment
Narrative	12	13	14	15	16	17	18	19	20	21	22
1	senate	treasurys	loan	oil	restaurant	pandemic	euros	mobile	software	google	film
2	voter	top list	lender	barrel	museum	patient	fiat	wireless	microsoft	facebook	disney
3	sen	year treasury	creditor	crude	ice	coronavirus	car maker	smartphone	app	amazon	studio
4	senator	nyse	borrower	natural gas	tree	dr	mainland	sprint	patent	social medium	song
5	obama	year note	tokyo	gasoline	christie	vaccine	nissan	broadband	iphone	tweet	youtube
6	ballot	biggest	subprime	crude oil	chicken	health care	volkswagen	mobile phone	pc	twitter	musical
7	nominee	bond price	foreclosure	million barrel	kitchen	covid	container	high speed	tech company	web site	viewer
8	bipartisan	percentage price	write andrew	opec	mall	cancer	euro euros	service provider	hacker	amazon com	music
9	caucus	list biggest	billion yen	metric	dining	medicare	suv	communications inc	microsoft corp	blog	videogame
10	pelosi	biggest percentage	mitsubishi	exxon	chef	fda	hk	internet service	windows	two people	olympics
11	majority leader	year bond	sovereign debt	barrel day	painting	public health	bmw	handset	apple inc	alphabet	walt
12	presidential candidate	yield year	nomura	oil company	beautiful	coronavirus pandemic	seng	cellular	computing	facebook inc	walt disney
13	cia	corporate bond	toyota motor	gallon	milk	irs	hong	long distance	operating system	jennifer	dvd
14	homeland security	benchmark year	motor corp	mercantile	gallery	medicaid	dax	verizon	software company	search engine	disney co
15	polling	investment grade	yen yen	mercantile exchange	bathroom	lockdown	kong	phone company	packard	google inc	broadcasting
16	senate majority	among common	subprime mortgage	brent	cream	clinical	billion euro	nortel	personal computer	write jennifer	inc top
17	president elect	write michael	current fiscal	refinery	cheese	physician	daimlerchrysler	federal communications	hewlett	dot com	sony corp
18	obama administration	treasury yield	repayment	drilling	borough	illness	sport utility	communications commission	hewlett packard	www	movie
19	presidential campaign	treasury bond	commercial bank	fossil	pizza	deduction	utility vehicle	lucent	machines	ventures	broadcast
20	midterm	year yield	collateral	grain	breakfast	prescription	hong kong	optic	business machines	york times	comedy
21	house senate	investors service	lending	york mercantile	chocolate	surgery	frankfurt	research motion	desktop	refund	sports
22	aide say	moody investors	refinance	mobil	fiction	nursing	hsbc holdings	fiber optic	intel corp	click	streaming
23	election day	yield rise	toyota	metric ton	renovation	income tax	dax index	nortel networks	computer system	database	documentary
24	minority leader	write david	fitch	countries	sandwich	tax credit	fall euro	aol	machines corp	com inc	game
25	turnout	bond issue	financial institution	oil future	coat	say dr	auto	motorola	computer maker	washington post	musician
26	centrist	yield fall	repay	katrina	office building	dr say	motor	internet access	jose calif	email	football
27	bernie sanders	yield move	bank also	heating	fish	therapy	tire	communications	palo	com say	cbs
28	sens	inversely	deposit	oil production	bike	food drug	automotive	time warner	palo alto	sad	league
29	candidacy	move inversely	bailout	nymex	garden	drug administration	car	subscriber	packard co	shipping	news corp
30	nonpartisan	bond market	balance sheet	oil field	flower	health official	hang seng	nokia	apple computer	web	theater
31	national committee	euro yen	austerity	oil market	clothe	disease control	vehicle	comcast	microsystems	card	concert
32	envoy	inversely price	lend	gasoline price	square	control prevention	seng index	grid	variant	fundraising	sony
33	write john	market action	honda	exporting	glass	centers disease	rise euro	worldcom	apple	advertiser	singer
34	house representatives	bond yield	say bank	organization petroleum	portrait	epidemic	truck	frontier	san jose	site	basketball
35	electoral	stock nasdaq	mortgage	exporting countries	shirt	medication	new model	cable	alto	chat	bowl
36	second term	bond	cent dollar	petroleum exporting	cow	care act	machinery	telecommunication company	oracle	ad	fan
37	nomination	high yield	big bank	oil producer	tall	affordable care	rubber	phone	computer	fake	actor
38	poll show	back security	pay down	matthew	fruit	tax return	spa	satellite	fbi	glitch	abc
39	presidential election	mortgage back	investment trust	crude price	rent	clinical trial	engine	customer service	hardware	online	actress
40	liberal	yield	sheet	barrel oil	menu	medical center	luxury	warner	software maker	mail	camera
41	election	moody	borrow	sweet crude	vegetable	prescription drug	diesel	warner inc	ebay	publication	fox
42	presidential	government bond	estate investment	light sweet	thick	health organization	hang	spectrum	laptop	advertise	roster
43	committee chairman	municipal	banking system	oil natural	festival	side effect	highway	frequency	intellectual property	classified	star
44	rep	basis point	borrowing	price oil	art	please include	ag	provider	server	advertisement	franchise
45	campaign say	junk	liquidity	commodity market	meal	disease	convertible	dial	printer	people use	audience
46	party leader	issuance	bankruptcy	cent barrel	grill	symptom	emission	telephone	high tech	anonymous	recording
47	party say	face value	distressed	comex division	bread	medicine	holdings ltd	footprint	intel	yahoo	dance
48	tax cut	bondholder	default	barrel new	deck	genetic	maker	turner	clara	los angeles	tooth
49	congressman	fix income	banking sector	say yesterday	cook	enrollment	suspension	bundle	gadget	angeles	console
50	ic	maturity	financial system	greenhouse	smell	clinic	drive	network	jose	kelly	visual

Continues

Meta Theme	College & Schooling	Others									
Narrative	23	24	25	26	27	28	29	30	31	32	33
1	student	dowjone	incorrectly	de	violence	goldman	flight	vaccination	coal	gm	gold
2	kid	dowjone com	version article	blasio	immigration	private equity	airline	egg	rise cent	sanders	steel
3	teacher	mutual	article incorrectly	de blasio	troop	sachs	plane	walker	tobacco	mart	metal
4	wine	etf	com corrections	pesos	terrorist	goldman sachs	airport	sound like	fall cent	wal	mining
5	high school	mutual fund	prosecutor say	class share	peace	climate change	boeing	co author	rand	wal mart	carbon
6	feel like	fund manager	chancellor	cola	taliban	equity firm	passenger	naturally	gain cent	chrysler	rail
7	baseball	ira	incorrectly say	coca	democracy	fargo	jet	prioritize	morris	ibm	aluminum
8	fun	fidelity	board member	mexico	soldier	wells fargo	airlines	president george	complete coverage	motor co	miner
9	public school	pension fund	racist	coca cola	state department	aig	cellphone	stroke	down cent	finance chief	european union
10	athlete	vanguard	conflict interest	journal editor	protester	enron	aviation	systemic	aapl	sedan	aerospace
11	flavor	bond fund	current former	newseditor wsj	gen	factset	sugar	global warming	lose cent	motors co	railway
12	classroom	trade fund	member say	newseditor	militant	sachs group	flu	etc	add cent	mart stores	imperial
13	laugh	fund say	front runner	write online	nations	investment banking	airbus	surely	drop cent	stores inc	gold price
14	stadium	stock fund	see corrections	editor newseditor	terrorism	morgan chase	cargo	virtue	cent new	motors corp	royal bank
15	exhibition	market fund	life say	ipc	rebel	wealth management	airways	narrative	shed cent	annual sale	toronto
16	list go	amy	corrections amplifications	grupo	united nations	citigroup inc	mayor say	parade	climb cent	store open	mining company
17	school student	equity fund	amplifications item	ipc index	diplomatic	accord research	earthquake	inevitably	decline issue	supply chain	group rise
18	funny	fund invest	former chairman	miller	diplomat	merger acquisition	kill people	precisely	up cent	max	copper
19	school say	fund company	amplifications	lion	refugee	jpm	northwest	supposedly	cigarette	auto industry	ounce
20	tell story	say fund	say person	banco	human right	street firm	delta	wisdom	cent cent	square foot	iron
21	girl	fund investor	conspiracy	del	foreign policy	security firm	federal fund	go wrong	philip	new car	mill
22	money see	fund rate	person say	city	al qaeda	accord thomson	see full	excerpt	cent	department store	mine
23	old man	investment fund	landlord	peso	commander	lynch co	boeing co	one could	five cent	auto maker	gas price
24	old son	taxable	one people	cement	treaty	salomon smith	write paul	good idea	two cent	motors	rio
25	college student	individual investor	include former	class	terror	familiar matter	lines	point view	beer	ford	ore
26	humor	pension	drone	beverage	nato	wells	air lines	op	diamond	company spokesman	canada
27	husband	retirement	step down	bellwether	killing	investment banker	infection	obvious	three cent	home depot	india
28	mom	tax free	heather	metro	embassy	hedge fund	vessel	somehow	four cent	rental	environmental
29	mother	investing	public relation	worth million	homeland	people familiar	cruise	columnist	africa	depot	steelmaker
30	coach	outflow	cfo	bottle	bombing	chase co	outbreak	one thing	group ltd	brand name	commodity price
31	young people	money market	former vice	real	security council	accord people	pilot	memoir	skid	apparel	brussels
32	marry	fund	former head	hunting	quarantine	underwriting	airplane	solve problem	sag	brands	locally
33	teenager	put money	senior executive	financial group	kremlin	matter say	missile	grasp	advancer	state control	comex
34	graduate	planner	ethic	share worth	extremist	hedge	port	perspective	consolidated	stores	forest
35	daughter	money manager	employee say	fix line	communist	person familiar	southwest	good thing	shed	bedroom	pan
36	bed	allocation	also call	se	nuclear weapon	citigroup	aboard	presumably	foods	lease	australia
37	inspiration	charity	respond request	surrender	condemn	financial firm	aircraft	sing	index shed	detroit	freight
38	boy	financial adviser	trump	ordinary	civil war	international group	temperature	universe	world big	battery	sub
39	teaching	funds	say former	share market	militia	stearns	evacuate	teenage	leap	chain	material
40	marriage	inflow	rape	broadcaster	terrorist attack	accord person	crew	observe	jewelry	industry analyst	resources
41	smile	buy stock	tenure	turnover	detention	merrill	accident	incumbent	outpace	brand	packaging
42	grow up	investments	say try	enterprises	riot	bear stearns	coastal	wake up	diversified	general motors	european
43	drink	investment company	chairwoman	soft	immediately respond	familiar	traveler	moral	american	grocery store	pacific
44	gay	institutional investor	karen	total million	humanitarian	leveraged	death toll	arguably	information technology	glut	rose
45	dream	portfolio	runner	worth billion	security force	stanley	fleet	correctly	edge up	hybrid	commodity
46	father	heir	injury	low income	citizenship	lynch	helicopter	ensue	decliner	time company	industries
47	love	raise money	succession	local	two country	morgan stanley	taxi	theory	board	bidder	group also
48	birthday	retiree	chief operating	go effect	insurgent	large bank	province	philosophy	shares	traction	rise price
49	school	payout	operating officer	also know	diplomacy	firm also	fly	honest	sink	outsourcing	month high
50	college	billion asset	staffer	fix	nuclear program	lehman	crash	brilliant	precious	cut cost	northern

OC Robustness Analysis and Extensions

This Appendix contains the discussions of robustness of our findings and extensions such as price informativeness interaction with exposure to selected individual narratives,

Sensitivity to recessions. The price informativeness can deteriorate in volatile market regimes and recessions, during which the stock prices can deviate from their fundamentals and long-terms levels. Thus, we check the sensitivity of our results by removing from the tests on narrative exposure and price informativeness the years with NBER recession periods (2001, 2008–2009, and 2020). The results in Table OC1 are very similar in coefficient magnitudes and significance to the analysis in the main text.

	One-year horizon				*Three-year horizon*			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
$\ln(M/A)_{n,t}$	0.023	0.032	0.033	0.032	0.048	0.061	0.059	0.062
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times \textit{Narrative Exposure}_{n,t}$	-0.016	-0.015	-0.015	-0.008	-0.029	-0.025	-0.025	-0.017
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\textit{Illiquidity}_{n,t}$	–	–	0.001	-0.001	–	–	0.000	-0.001
			(0.211)	(0.011)			(0.924)	(0.389)
$\textit{MAX}_{n,t}$	–	–	0.001	0.009	–	–	0.014	0.011
			(0.599)	(0.001)			(0.129)	(0.001)
$\textit{DOB}_{n,t}$	–	–	–	-0.009	–	–	–	-0.008
				(0.001)				(0.001)
$\textit{Inst. Ownership}_{n,t}$, %	–	–	–	0.003	–	–	–	0.004
				(0.001)				(0.021)
R^2 (%)	77.94	79.40	79.46	77.54	57.04	60.31	60.50	55.28
Obs.	3,151	3,151	3,151	946	2,470	2,470	2,470	859
Factor betas	–	FF4	FF4	FF4	–	FF4	FF4	FF4
Fundamentals	–	Yes	Yes	Yes	–	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High Average Exposure	-0.009	-0.009	-0.009	-0.004	-0.018	-0.008	-0.006	-0.005
	(0.001)	(0.001)	(0.001)	(0.016)	(0.248)	(0.330)	(0.358)	(0.240)

Table OC1: Price Informativeness and Narrative Exposure in Recessions-free Periods. The table shows aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its interaction with $\textit{Narrative Exposure}_{n,t}$ defined in (3). The model is estimated as the two-stage regression (5) for one- and three-year horizons. The sample period is from 1998 to 2021, with annual frequency, and years with NBER recession periods (2001, 2008–2009, and 2020) removed from estimation. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and market value $\ln(M/A)$ is standardized to unit standard deviation. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and the number of observations (Obs.) are average numbers from the cross-sectional stage.

Narrative exposure changes and price informativeness. To test the sensitivity of the quasi-causal analysis provided in Table 7, we repeat the computations using as the threshold for treated firms 20% and 30% change in rank from one year to the next. The results provided in Table OC2 show that qualitatively the negative effect on price informativeness is similar across different thresholds. The significance of the results deteriorates slightly in Panel B for the one-year horizon, though they remain significant at the 10% significance level.

	One-year horizon				*Three-year horizon*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 20% Rank change								
$\ln(M/A)_{n,t}$	0.0095	0.0105	0.0103	0.0102	0.0159	0.0206	0.0206	0.0207
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(M/A)_{n,t} \times$ *Treated*	-0.0035	-0.0043	-0.0040	-0.0038	-0.0084	-0.0077	-0.0073	-0.0074
	(0.036)	(0.004)	(0.012)	(0.022)	(0.001)	(0.003)	(0.004)	(0.005)
R^2 (%)	72.15	72.52	72.70	73.39	46.60	48.11	48.38	49.60
Obs.	51,732	51,732	51,732	51,732	39,222	39,222	39,222	39,222
Panel B: 30% Rank change								
$\ln(M/A)_{n,t}$	0.0123	0.0120	0.0120	0.0120	0.0215	0.0244	0.0247	0.0250
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(M/A)_{n,t} \times$ *Treated*	-0.0042	-0.0044	-0.0040	-0.0040	-0.0146	-0.0138	-0.0133	-0.0137
	(0.069)	(0.037)	(0.057)	(0.058)	(0.000)	(0.000)	(0.000)	(0.000)
R^2 (%)	68.48	68.94	69.12	69.90	43.00	44.39	44.67	46.26
Obs.	22,260	22,260	22,260	22,260	16,434	16,434	16,434	16,434
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	No	Yes	No	Yes	No	Yes	No
Year FE	No	Yes	Yes	No	No	Yes	Yes	No
Sector x Year FE	No	No	No	Yes	No	No	No	Yes

Table OC2: Narrative Exposure Changes and Price Informativeness.
The table shows the aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and the change in price informativeness for firms (coefficient for interaction with *Treated*) that experienced a large (20% and 30% rank change) annual change in the narrative exposure relative to comparable firms. The model (6) is estimated as the absorbing least squares for one- and three-year horizons (columns 1-4 and 5-8, respectively). Controls include four-factor betas, fundamental and stock characteristics, and various fixed effects. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and market value $\ln(M/A)$ is standardized to unit standard deviation. $p-$values for coefficients in parentheses use standard errors clustered at year-firm level.

Individual narratives and price informativeness. For the ease of exposition, we also group most of the originally recovered 33 narratives into a smaller set of 12 narratives, based on the similarity of their top terms to broader themes by summing $\theta_{l,\tau}$ across narratives for each sub-group on each day. Table OB2 in the Online Appendix shows the top terms associated with individual narratives and the grouping of the latter into 12 topical narratives based on manual classification of the top-100 representative unigrams and bigrams. We abbreviate these topical narratives for the tables as follows: POLY: Politics, REGL: Regulation, MCRO: Macroeconomy, EQTY: Equity markets, FINC: Fixed income markets, ENGY: Energy markets, STPL: Consumer staples, HLTH: Healthcare, AUTO: Automotive, TLCO: Telecommunications and social media, ENTM: Entertainment, SCHL: College and schooling.

Figure OC1 depicts the evolution of the attention level devoted to the identified topical narratives, following Eq. (1). There is substantial variation in the level of attention devoted to each narrative in the WSJ, reflecting the concept that the news media tends to focus on different narratives at different times, due to changing economic and political conditions, and the changing interests and sentiments of market participants. For instance, the "Equity markets" narrative accounted for a sizeable chunk of the WSJ's attention allocation in the early sample period, but declined over time, while attention to "Regulation" and "Political" narratives grew. Overall, the evident changes in attention allocation to different narratives could impact agents' perspectives regarding the prospects of individual assets, resulting in trading decisions that may or may not distort prices.

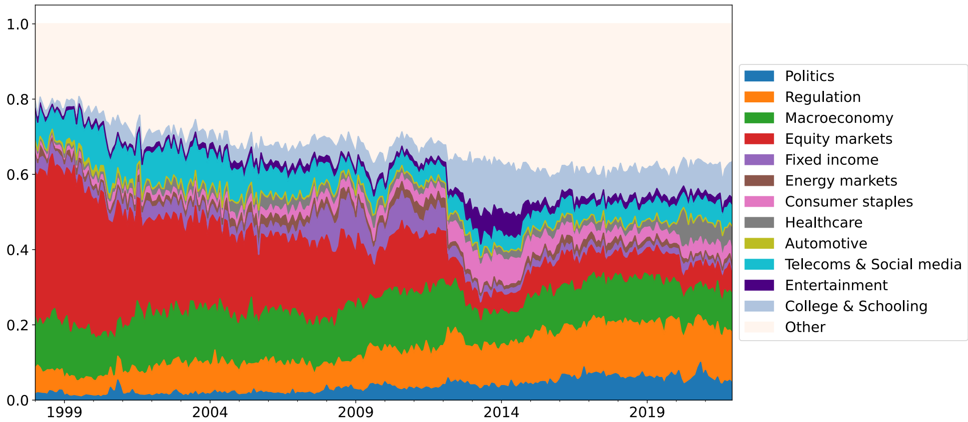

Figure OC1: Evolution of Narrative Attention. The figure shows the evolution of attention, from Eq. (1), dedicated to the identified narratives over time after grouping them into 12 themes.

Table OC3 shows in Panel A that attention to different groups of narratives often moves in opposite directions (partially, by construction), and one can potentially identify more precise clusters of topics that inspire interest of media at the same time. Exposure to narratives (Panel B), however, is always positively correlated across different firms, so that exposure to any topic can reflect the general sensitivity of a stock price to media talks, which, as we have seen in the main part of the paper, renders prices uninformative about future fundamentals.

While in the main text we concentrated on the aggregate narrative exposure, now we analyze whether exposure to individual narrative groups affects the price informativeness. Recall from the main text that our main model is specified as the Fama and MacBeth (1973) regression of future earnings h years from today relative to current assets, $E_{n,t+h}/A_{n,t}$, on current earnings, market value relative to assets, $\ln(M_{n,t}/A_{n,t})$, the interaction of market value and particular narrative exposure, and controls:

$$\frac{E_{n,t+h}}{A_{n,t}} = a + b_{0,h}\frac{E_{n,t}}{A_{n,t}} + b_{1,h}\ln\frac{M_{n,t}}{A_{n,t}} + b_{2,h}\ln\frac{M_{n,t}}{A_{n,t}} \times |\beta^{narr}_{n,t}| + b^{\top}_{x,h}X_{n,t} + \varepsilon_{n,t+h}, \qquad \text{(OC1)}$$

where h is one or three years, and $|\beta^{narr}_{n,t}|$ denotes the narrative exposure of firm n at time t with respect to a particular narrative group. The vector of controls, $X_{n,t}$, includes the narrative exposure used in interaction term, four-factor model betas, fundamental variables $\ln(Assets)$,

	POLY	REGL	MCRO	EQTY	FINC	ENGY	STPL	HLTH	AUTO	TLCO	ENTM	SCHL
	Panel A: Narrative Attention.											
POLY	1.000	0.551	-0.180	-0.548	-0.256	-0.080	0.137	0.147	-0.298	-0.369	0.070	0.271
REGL	0.551	1.000	-0.215	-0.606	-0.235	-0.013	0.116	0.204	-0.372	-0.307	0.094	0.218
MCRO	-0.180	-0.215	1.000	0.077	0.294	0.184	-0.298	-0.131	-0.022	-0.113	-0.314	-0.334
EQTY	-0.548	-0.606	0.077	1.000	0.218	0.008	-0.493	-0.329	0.497	0.375	-0.353	-0.658
FINC	-0.256	-0.235	0.294	0.218	1.000	0.142	-0.217	-0.158	0.069	-0.113	-0.218	-0.335
ENGY	-0.080	-0.013	0.184	0.008	0.142	1.000	-0.105	-0.084	-0.045	-0.129	-0.136	-0.215
STPL	0.137	0.116	-0.298	-0.493	-0.217	-0.105	1.000	0.059	-0.229	-0.277	0.547	0.691
HLTH	0.147	0.204	-0.131	-0.329	-0.158	-0.084	0.059	1.000	-0.217	-0.057	-0.008	0.197
AUTO	-0.298	-0.372	-0.022	0.497	0.069	-0.045	-0.229	-0.217	1.000	0.203	-0.187	-0.302
TLCO	-0.369	-0.307	-0.113	0.375	-0.113	-0.129	-0.277	-0.057	0.203	1.000	-0.154	-0.290
ENTM	0.070	0.094	-0.314	-0.353	-0.218	-0.136	0.547	-0.008	-0.187	-0.154	1.000	0.577
SCHL	0.271	0.218	-0.334	-0.658	-0.335	-0.215	0.691	0.197	-0.302	-0.290	0.577	1.000
	Panel B: Narrative Exposure.											
POLY	1.000	0.424	0.362	0.221	0.258	0.381	0.352	0.443	0.228	0.275	0.421	0.415
REGL	0.424	1.000	0.407	0.283	0.312	0.395	0.376	0.417	0.275	0.338	0.423	0.405
MCRO	0.362	0.407	1.000	0.341	0.381	0.415	0.355	0.373	0.342	0.393	0.397	0.381
EQTY	0.221	0.283	0.341	1.000	0.397	0.336	0.261	0.203	0.422	0.377	0.269	0.263
FINC	0.258	0.312	0.381	0.397	1.000	0.355	0.307	0.263	0.378	0.334	0.319	0.283
ENGY	0.381	0.395	0.415	0.336	0.355	1.000	0.309	0.375	0.324	0.352	0.398	0.331
STPL	0.352	0.376	0.355	0.261	0.307	0.309	1.000	0.347	0.307	0.296	0.378	0.530
HLTH	0.443	0.417	0.373	0.203	0.263	0.375	0.347	1.000	0.204	0.297	0.438	0.402
AUTO	0.228	0.275	0.342	0.422	0.378	0.324	0.307	0.204	1.000	0.370	0.277	0.285
TLCO	0.275	0.338	0.393	0.377	0.334	0.352	0.296	0.297	0.370	1.000	0.330	0.298
ENTM	0.421	0.423	0.397	0.269	0.319	0.398	0.378	0.438	0.277	0.330	1.000	0.411
SCHL	0.415	0.405	0.381	0.263	0.283	0.331	0.530	0.402	0.285	0.298	0.411	1.000

Table OC3: Correlations of Attention Levels and Exposure to Narrative Groups. The table shows the correlations among narrative attention levels (Panel A) and among narrative exposure levels (Panel B), the latter computed from the firm-year panel data. Each year, the narrative exposure levels are winsorized at 5% and 95%.

Debt/Assets, *Cash/Assets*, *Ppent/Assets*, *Capex/Assets*, *Sales/Assets*, *R&D/Assets*, and economic sector dummies (eight one-digit SIC codes after excluding the financial sector). All continuous variables are winsorized at 5% and 95% for each year in the sample period. The market value variable $\ln(M/A)$ is standardized to unit variance each year in the cross-section so that the coefficient, $b_{1,h}$, directly provides the proxy for price informativeness following Bai, Philippon, and Savov (2016). The coefficient $b_{2,h}$, therefore, reveals how price informativeness interacts with a particular narrative exposure.

The results in Table OC4 clearly show that price informativeness significantly decreases for stocks with high narrative exposure for both the one- and three-year future horizons and for all narratives (except for SCHL for the three-year horizon). The pattern does not seem to be dependent upon the perceived relevance of the specific narratives to certain economic fundamentals or industries. This result delivers a profound message: firms whose stock prices co-vary substantially with media narratives, in general, tend to absorb irrelevant information that renders prices uninformative. At the end of each panel in Table OC4, we estimate the marginal change in the incremental price informativeness, i.e., the interaction term, conditional on periods of high attention level to a particular narrative. For this, we regress the time-

	POLY	REGL	MCRO	EQTY	FINC	ENGY	STPL	HLTH	AUTO	TLCO	ENTM	SCHL
Panel A: One-year horizon.												
$\ln(M/A)_{n,t}$	0.010	0.010	0.010	0.010	0.011	0.011	0.010	0.011	0.010	0.010	0.010	0.011
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times \lvert\beta^{narr}_{n,t}\rvert$	-0.074	-0.121	-0.150	-0.188	-0.064	-0.052	-0.064	-0.059	-0.015	-0.098	-0.047	-0.107
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	79.08	79.08	79.08	79.07	79.10	79.10	79.09	79.09	79.09	79.07	79.08	79.09
Obs.	3,151	3,151	3,151	3,151	3,151	3,151	3,151	3,151	3,151	3,151	3,151	3,151
High Attention	-0.067	-0.041	-0.006	-0.114	-0.030	-0.028	-0.027	-0.050	-0.008	-0.051	-0.038	-0.061
Marginal Effect	(0.001)	(0.065)	(0.854)	(0.010)	(0.071)	(0.011)	(0.206)	(0.042)	(0.083)	(0.001)	(0.008)	(0.004)
Panel B: Three-year horizon.												
$\ln(M/A)_{n,t}$	0.023	0.026	0.027	0.025	0.028	0.027	0.028	0.025	0.029	0.019	0.025	0.012
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.347)
$\ln(M/A)_{n,t} \times \lvert\beta^{narr}_{n,t}\rvert$	-0.104	-0.204	-0.303	-0.323	-0.114	-0.095	-0.142	-0.107	-0.026	-0.139	-0.082	-0.093
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.310)
R^2 (%)	59.89	59.78	59.76	59.89	59.83	59.92	59.87	59.73	59.89	59.76	59.73	59.96
Obs.	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470
High Attention	-0.048	-0.102	0.125	-0.114	-0.048	-0.046	-0.152	-0.085	0.001	-0.118	-0.075	0.127
Marginal Effect	(0.453)	(0.078)	(0.197)	(0.179)	(0.099)	(0.053)	(0.093)	(0.001)	(0.864)	(0.027)	(0.001)	(0.548)

Table OC4: Price Informativeness and Narrative Exposure.
The table shows aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its interaction with exposure to selected narrative groups $|\beta^{narr}_{n,t}|$. The model is estimated as the two-stage regression (5) for one- and three-year horizons (Panels A and B, respectively). Below each panel, the mean interaction term coefficient is computed, conditional on high (above the mean) attention to a narrative. Controls include four-factor betas, fundamental variables, and sector dummies. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and market value $\ln(M/A)$ is standardized to unit standard deviation. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. $R^2(\%)$ and the number of observations (Obs.) are average numbers from the cross-sectional stage.

series of the interaction term coefficient $b_{2,h}$ from the cross-sectional stage of the Fama-MacBeth procedure on a constant and a dummy variable that equals one for the years of high attention to the specific narrative, defined as periods when attention to the narrative is above its sample mean, and zero otherwise. We report the coefficient on the dummy variable along with its $p-$value. For the majority of narratives for the one-year horizon and for five out of 12 narratives for the three-year horizon, high attention significantly (at 5% level) exacerbates the loss of price informativeness for exposed stocks. Which narratives have a stronger marginal effect is hardly anticipated ex-ante—e.g., the Macroeconomy (MCRO) narrative is insignificant, while Entertainment (ENTM) and Telecoms & Social Media (TLCO) are both significant.

To illustrate the economic magnitude of these effects, we standardize the absolute narrative betas each year in the panel data.[19] For the one-year horizon, the absolute exposure to individual narratives significantly decreases price informativeness by almost identical magnitudes (-0.006 to -0.007) for a standard deviation increase in the exposures. For the three-year horizon, we obtain slightly more heterogeneity in economic magnitudes but still find an almost uniform significance of interaction term coefficients, with the exception of the SCHL narrative.

[19] Full results are available upon request.